

---

# **MÉTHODES D'ÉVALUATION DES PROGRAMMES :**

**Mesure et attribution des résultats des programmes**

**Troisième édition**

**Pratiques d'examen et études  
Revue gouvernementale et services de qualité  
Direction du sous-contrôleur général  
Secrétariat du Conseil du Trésor du Canada**

---

---

Also available in English under the title :

*Program Evaluation Methods : Measurement and  
Attribution of Program Results*

© Ministre des Travaux publics et des Services gouvernementaux  
Publié par la Direction des affaires publiques  
Secrétariat du Conseil du Trésor du Canada

---

---

# Table des matières

	Page
<b>Chapitre 1 INTRODUCTION</b>	<b>1</b>
1.1 Objectifs et plan	2
1.2 Processus d'évaluation	2
1.3 Questions à évaluer	2
<b>Chapitre 2 STRATÉGIES D'ÉVALUATION</b>	<b>13</b>
2.1 Inférence causale en évaluation	13
2.2 Inférences causales	14
2.3 Stratégies d'évaluation	20
2.4 Évaluations crédibles	23
2.4.1 Critères de recherche .....	23
2.4.2 Critères du contexte décisionnel.....	28
2.4.3 Stratégies multiples .....	34
2.5 Résumé	36
<b>Chapitre 3 MODÈLES D'ÉVALUATION</b>	<b>37</b>
3.1 Introduction	37
3.2 Modèles expérimentaux aléatoires	41
3.3 Modèles quasi expérimentaux	47
3.3.1 Modèles dont les mesures sont prises avant et après le programme .....	47
3.3.2 Modèles à série temporelle ou modèles chronologiques.....	49
3.3.3 Modèles dont les mesures sont prises après le programme seulement..	51
3.4 Modèles implicites	54
3.5 Utilisation des modèles de causalité pour l'évaluation	58
3.6 Résumé	61
<b>Chapitre 4 MÉTHODES DE COLLECTE DES DONNÉES</b>	<b>62</b>
4.1 Introduction	62

---

<b>4.2</b>	<b>Dépouillement de la documentation spécialisée</b>	<b>65</b>
<b>4.3</b>	<b>Étude de dossiers</b>	<b>68</b>
<b>4.4</b>	<b>Observations directes</b>	<b>72</b>
<b>4.5</b>	<b>Enquêtes</b>	<b>75</b>
<b>4.6</b>	<b>Consultation de spécialistes</b>	<b>79</b>
<b>4.7</b>	<b>Études de cas</b>	<b>82</b>
<b>4.8</b>	<b>Résumé</b>	<b>85</b>
<b>Chapitre 5</b>	<b>MÉTHODES ANALYTIQUES</b>	<b>86</b>
<b>5.1</b>	<b>Introduction</b>	<b>86</b>
<b>5.2</b>	<b>Analyse statistique</b>	<b>86</b>
<b>5.3</b>	<b>Analyse de l'information qualitative</b>	<b>96</b>
<b>5.4</b>	<b>Analyse des autres résultats des programmes</b>	<b>99</b>
<b>5.5</b>	<b>Utilisation de modèles</b>	<b>102</b>
5.5.1	Modèles de simulation.....	102
5.5.2	Modèles d'entrées-sorties.....	105
5.5.3	Modèles micro-économiques .....	107
5.5.4	Modèles macro-économiques.....	110
5.5.5	Modèles statistiques .....	112
<b>5.6</b>	<b>Analyse coûts-avantages et analyse coût-efficacité</b>	<b>114</b>
<b>5.7</b>	<b>Résumé</b>	<b>121</b>
<b>Chapitre 6</b>	<b>CONCLUSIONS</b>	<b>122</b>
<b>Annexe 1</b>	<b>ENQUÊTES</b>	<b>123</b>
<b>1.1</b>	<b>Échantillonnage</b>	<b>123</b>
<b>1.2</b>	<b>Méthodes d'enquête</b>	<b>125</b>
<b>1.3</b>	<b>Instruments de mesure</b>	<b>129</b>
<b>1.4</b>	<b>Estimation des coûts</b>	<b>131</b>
<b>1.5</b>	<b>Avantages et inconvénients</b>	<b>131</b>
<b>Annexe 2</b>	<b>GLOSSAIRE</b>	<b>134</b>

---

<b>Annexe 3 BIBLIOGRAPHIE</b>	<b>142</b>
<b>Annexe 4 AUTRES RÉFÉRENCES</b>	<b>155</b>
<b>Figures et tableaux</b>	
<b>Figure 1 — Le processus d'évaluation</b>	<b>3</b>
<b>Figure 2 — Les modèles expérimentaux produisent la meilleure estimation des effets incrémentiels</b>	<b>16</b>
<b>Figure 3 — À chaque aspect sa stratégie</b>	<b>21</b>
<b>Figure 4 — Modèle de l'effet d'une taxe d'accise</b>	<b>108</b>
<b>Tableau 1 — Aspects fondamentaux de l'évaluation d'un programme</b>	<b>5</b>
<b>Tableau 2 — Éléments nécessaires à la crédibilité des évaluations</b>	<b>23</b>
<b>Tableau 3 — Exemple de statistiques descriptives</b>	<b>88</b>
<b>Tableau 4 — Autres données descriptives</b>	<b>91</b>
<b>Tableau 5 — Exemple de statistiques descriptives</b>	<b>93</b>

## Chapitre 1

### INTRODUCTION

L'évaluation du rendement des programmes est un élément clé de la stratégie de gestion des résultats du gouvernement fédéral. Le cycle de chaque programme (conception, mise en œuvre et évaluation) s'inscrit dans celui, plus large, du Système de gestion des dépenses du gouvernement. Les plans établissent les objectifs du programme et les critères d'évaluation de son succès; les rapports sur le rendement évaluent les résultats.

La mesure du rendement est un élément essentiel de ce cycle. Les évaluations devraient fournir au moment voulu des constatations et des conclusions à la fois pertinentes, fiables et objectives sur le rendement du programme, grâce à des méthodes valides et fiables de collecte et d'analyse des données. Idéalement, les évaluations devraient présenter les constatations et les conclusions de façon claire et équilibrée, tout en précisant leur degré de fiabilité.

Ce manuel est un exposé des méthodes appropriées pour atteindre les objectifs analytiques. Dans une large mesure, les défis de l'évaluation sont typiques de ceux qui se posent dans toute la recherche en sciences sociales. La documentation pertinente foisonne d'excellentes descriptions du bon et du mauvais usage des méthodes d'évaluation. À cet égard, il convient de souligner que la documentation sur les techniques et les objets de la recherche en sciences sociales traite de façon beaucoup plus détaillée que nous le faisons des questions méthodologiques abordées dans ces pages. Il faut aussi souligner que **bien peu des méthodes dont nous parlons ici devraient être utilisées sans que l'on consulte d'autres ouvrages et publications ou des praticiens d'expérience**. C'est pour cette raison que la plupart des rubriques de notre manuel comprennent une liste de tels documents.

## 1.1 Objectifs et plan

Il est généralement difficile de choisir les méthodes appropriées d'évaluation d'un programme. C'est d'autant plus complexe en raison des nombreux points qui peuvent devoir faire l'objet de l'évaluation, du grand nombre de méthodes susceptibles d'être utilisées pour recueillir et analyser les données, compte tenu des ressources et du temps disponibles, ainsi que de la nécessité de faire en sorte que tous les points pertinents soient analysés.

Ce manuel a été conçu pour aider les praticiens et les autres parties intéressées à comprendre les facteurs méthodologiques intervenant dans la mesure et l'évaluation des résultats des programmes. L'accent est mis sur les avantages et les inconvénients des diverses méthodes étudiées. Il ne faut pas considérer le manuel comme un ensemble de lignes directrices contenant des instructions détaillées pour les évaluateurs, mais plutôt comme un exposé des éléments méthodologiques dont il faut tenir compte pour réaliser une étude d'évaluation crédible des résultats d'un programme.

## 1.2 Processus d'évaluation

Les évaluations comprennent trois phases (représentées à la figure 1) :

- l'étude préparatoire à l'évaluation ou le cadre (la planification);
- l'étude d'évaluation;
- la prise de décisions fondées sur les constatations et les recommandations.

L'étude préparatoire à l'évaluation dégage les principaux points et les principales questions sur lesquels l'évaluation doit porter, en déterminant les méthodes les mieux adaptées à la collecte des données recherchées. L'information est ensuite présentée au client sous forme d'options d'évaluation, pour qu'il choisisse l'approche qui lui convient. L'étude d'évaluation elle-même ne peut commencer qu'une fois son mandat établi. Les données sont alors recueillies et analysées de façon qu'on puisse en dégager des constatations sur les points à évaluer (voir les sous-études 1, 2 et 3 à la figure 1). On se base alors sur les constatations et sur les recommandations pour prendre des décisions sur l'avenir du programme. La discussion des constatations permet d'aider à assurer qu'on rende compte des résultats.

## 1.3 Questions à évaluer

Lorsqu'on envisage les questions à évaluer et les méthodes à utiliser à cet égard, il faut généralement faire une distinction entre deux types de résultats :

- **les extrants opérationnels;**

Figure 1

**Le processus d'évaluation**

## A) Étude préparatoire

Familiarisation  
*(Profil du programme et  
examen documentaire)*



Analyse de la raison d'être  
*(Modèle logique)*



Formulation et analyse des  
options d'étude d'évaluation  
*(Sujets, méthodes et coûts)*



## B) Évaluation

Schéma d'évaluation pour l'option retenue

Sous-étude 1

Sous-étude 2

Sous-étude 3



Synthèse des conclusions



Recommandations



- **les résultats** qui comprennent à la fois des avantages pour les clients du programme (et parfois aussi les inconvénients imprévus pour les clients et pour d'autres personnes) ainsi que les retombées connexes, associées aux objectifs du programme (par exemple, la création d'emplois, l'amélioration de la santé, de la sécurité et du mieux-être, ainsi que la sécurité nationale).

Les évaluations portent habituellement sur de nombreuses questions. Bien sûr, chaque programme est unique, mais les questions à évaluer peuvent souvent être groupées sous les trois rubriques suivantes.

- **Maintien de la pertinence** : Mesure dans laquelle le programme conserve sa pertinence pour les priorités de l'État et les besoins des citoyens.
- **Résultats** : Mesure dans laquelle le programme atteint ses objectifs dans les limites du budget, sans avoir d'importantes retombées indésirables.
- **Rentabilité** : Mesure dans laquelle le programme fait appel à la méthode la plus appropriée, la plus efficiente et la plus économique pour atteindre ses objectifs.

Tableau 1

### Aspects fondamentaux de l'évaluation d'un programme

#### A. MAINTIEN DE LA PERTINENCE

##### Raison d'être du programme

- Dans quelle mesure les objectifs et le mandat du programme sont-ils encore pertinents?
- Les activités et les extrants opérationnels sont-ils compatibles avec le mandat du programme et liés de façon plausible aux objectifs et aux autres résultats escomptés?

#### B. RÉSULTATS

##### Atteinte des objectifs

- De quelle manière et dans quelle mesure le programme a-t-il permis d'atteindre des objectifs appropriés?

##### Effets et conséquences

- Quels avantages pour la clientèle et quelles retombées, volontaires ou involontaires, ont résulté de l'exécution du programme?
- De quelle manière et dans quelle mesure peut-on dire que le programme complète, reprend, chevauche ou contredit d'autres programmes?

#### C. RENTABILITÉ

##### Évaluation des solutions de remplacement

- Existe-t-il d'autres méthodes plus rentables pour atteindre les objectifs et obtenir les résultats escomptés?
- Existe-t-il des moyens plus rentables d'exécuter le programme actuel?

**En ce qui concerne les méthodes d'évaluation, on distingue deux groupes de questions à évaluer.** Le premier est celui des questions liées à la théorie et à la structure du programme, c'est-à-dire à sa raison d'être et les autres solutions envisageables. Prenons par exemple un programme d'aide à l'industrie dans le cadre duquel le gouvernement accorde des subventions ponctuelles à tel ou tel projet. Dans ce cas, la question fondamentale qu'il faut se poser sur la raison d'être du programme est la suivante : «Pourquoi le gouvernement inciterait-il des entreprises à lancer des projets qu'elles n'entreprendraient pas autrement?» Pour que le programme satisfasse à ce critère, il doit exister une justification convaincante dans la politique publique. Les avantages sociaux du programme pour le Canada doivent l'emporter sur ses coûts

sociaux afin qu'il puisse en valoir la peine, même si son rendement n'est pas suffisamment élevé pour qu'une entreprise soit disposée à investir ses propres fonds. Cela pourrait arriver parce que le gouvernement peut répartir son risque sur un grand nombre de projets qui, pris individuellement, représenteraient un trop gros risque pour une entreprise privée.

Pour mieux comprendre cette question, prenons un deuxième exemple, celui d'un programme spécial d'enseignement du français ou de l'anglais aux immigrants. Sa raison d'être pourrait être justifiée par des lacunes des systèmes scolaires actuels. On se demanderait donc pourquoi le gouvernement fédéral devrait offrir un tel programme : parce que les écoles sont surpeuplées, ou parce que seules les écoles privées en offrent un, mais à un prix inabordable pour beaucoup d'immigrants? On pourrait aussi juger nécessaire d'offrir aux immigrants plus de cours d'anglais, tout en concluant qu'une aide directe aux écoles qui en offrent déjà pourrait être une solution de remplacement plus efficace.

L'autre groupe de questions à évaluer (atteinte des objectifs, effets et conséquences) est lié aux résultats du programme. Qu'est-il arrivé grâce à celui-ci? Revenons à notre premier exemple et supposons que le gouvernement a octroyé une subvention pour un projet qui a créé dix emplois. Peut-on dire, en ce qui concerne son objectif de création d'emplois sous-jacent, que le programme a été fructueux pour cette raison? Avant de faire une déclaration crédible à ce sujet, nous devons répondre aux questions suivantes :

- Le projet aurait-il été entrepris sans l'aide de l'État? Si oui, aurait-il été réalisé à plus petite échelle?
- Les personnes embauchées étaient-elles en chômage, ou ont-elles simplement quitté un autre emploi? Si les emplois qu'elles occupaient sont restés vacants, ou s'ils ont été simplement repris par des personnes qui avaient déjà du travail, le projet n'a peut-être eu aucun effet net de création d'emplois. Dans ce cas, l'objectif sous-jacent n'aurait pas été atteint.

L'évaluation doit porter aussi bien sur les effets attendus qu'imprévus du programme. Dans l'exemple qui nous intéresse, les effets attendus pourraient être une augmentation du revenu personnel ou l'accroissement des exportations canadiennes. Les effets imprévus pourraient comprendre un soutien financier accru des entreprises étrangères, aux dépens des sociétés canadiennes, ou encore le maintien d'activités incompatibles avec la restructuration indispensable du secteur industriel visé. Si le projet avait pu être réalisé sans l'aide de l'État, il est impossible d'imputer au programme d'aide le blâme de ses effets néfastes, pas plus que de lui attribuer le mérite de ses effets favorables.

Par ailleurs, le programme de notre deuxième exemple pourrait avoir pour objectif l'amélioration de l'aptitude à la lecture des immigrants participants, mais il pourrait aussi avoir d'autres retombées, comme le revenu auquel des immigrants devraient renoncer pour assister aux cours, les emplois ou le revenu accru qu'ils pourraient obtenir grâce à leur apprentissage de l'anglais (s'il ne s'agit pas là d'objectifs du programme) et les répercussions du programme sur les écoles qui offrent des cours analogues, par exemple, une baisse de leur clientèle ou la mise à pied d'enseignants.

Dans le tableau 1, les questions à évaluer sont groupées sous deux rubriques : l'une théorique (raison d'être et solutions de remplacement) et l'autre concrète, axée sur les résultats du programme (atteinte des objectifs, effets et conséquences). La deuxième rubrique englobe deux grands types de problèmes d'analyse, à savoir a) *les problèmes de mesure*, sur la façon de mesurer les résultats d'un programme et b) *les problèmes d'attribution*, sur la manière de déterminer si les résultats constatés sont attribuables au programme (et dans quelle mesure ils le sont). Ces deux problèmes et la façon d'utiliser divers outils méthodologiques pour les surmonter constituent le thème principal du manuel.

Il faudrait toutefois souligner que bon nombre de questions méthodologiques liées à la détermination des résultats d'un programme sont tout aussi indissociables de l'analyse de sa raison d'être et des autres solutions envisageables. Par exemple, si l'on remet en question le maintien d'un programme, on peut procéder à une analyse approfondie pour en mesurer la pertinence (Poister, 1978, p. 6 et 7; Kamis, 1979). En pareil cas, il peut se poser des problèmes de mesure semblables à ceux qui se manifestent lorsqu'on étudie les résultats d'un programme.

Néanmoins, l'analyse de ces résultats pose au moins un problème qui ne se présente pas lorsqu'on étudie les aspects théoriques d'un programme, celui de l'attribution. C'est typiquement le plus ardu des aspects de l'évaluation, et pourtant le plus important. Les difficultés d'attribution sont l'un des principaux sujets traités dans ces pages.

Après avoir souligné les difficultés associées à l'attribution des résultats d'un programme, nous devrions préciser que leur ampleur varie selon le genre de programme et les résultats étudiés. Ainsi, la satisfaction de la clientèle pourrait être le résultat attendu d'un programme de service, lequel peut alors être le seul facteur plausible auquel on puisse attribuer le niveau de satisfaction observé. Un modèle d'évaluation relativement rudimentaire fondé sur quelques arguments seulement peut alors suffire à attribuer les résultats. Il reste toutefois que l'attribution demeure un aspect à traiter soigneusement, puisqu'un lien apparemment manifeste avec le programme peut se révéler invalide. Par exemple, le mécontentement à l'égard des Centres d'emploi du Canada peut refléter les conditions économiques globales, plutôt qu'être attribuable au service effectivement assuré grâce au programme. Dans ce cas, la détermination du niveau de satisfaction de la clientèle attribuable au programme lui-même pourrait se révéler très difficile.

Enfin, l'évaluation ne devrait pas considérer le programme comme une «boîte noire» transformant automatiquement des intrants en extrants et en effets. Cette approche laisse une énorme zone grise dans notre compréhension des raisons du succès ou de l'échec des programmes. Pour interpréter une constatation quelconque sur les résultats d'un programme, il faut être en mesure de déterminer si la réussite (ou l'échec) est attribuable au succès (ou à l'échec) du cadre théorique du programme, de sa mise en œuvre ou de ces deux éléments. Pour faire une interprétation comme celle-là — élément indispensable à la formulation de recommandations en vue de la prise de décisions —, il faut connaître la dynamique générale et les extrants opérationnels du programme. C'est cette connaissance qui permet à l'évaluateur d'analyser les extrants dans le contexte de la raison d'être du programme et de son cadre théorique afin de déterminer les raisons de sa réussite ou de son échec.

## Références : Introduction à l'évaluation

Alberta, ministère du Trésor, *Measuring Performance : A Reference Guide*, Edmonton, septembre 1996.

Alkin, M.C., *A Guide for Evaluation Decision Makers*, Thousand Oaks : Sage Publications, 1986.

Berk, Richard A. et Peter H. Rossi, *Thinking About Program Evaluation*, Thousand Oaks : Sage Publications, 1990.

Canada, Secrétariat du Conseil du Trésor, *Aborder les années 90 : Perspectives gouvernementales pour l'évaluation de programmes*, Ottawa, 1991.

Canada, Secrétariat du Conseil du Trésor, «Examen, vérification interne et évaluation», *Manuel du Conseil du Trésor*, Ottawa, 1994.

Canada, Secrétariat du Conseil du Trésor, *Guide de la gestion de la qualité*, Ottawa, octobre 1992.

Canada, Secrétariat du Conseil du Trésor, *Guides des services de qualité : Services de qualité — Tour d'horizon*, Ottawa, octobre 1995;

*Guide I — Consultation des clients*, Ottawa, octobre 1995;

*Guide II — Mesure de la satisfaction des clients*, Ottawa, octobre 1995;

*Guide III — Collaboration avec les syndicats*, Ottawa, octobre 1995;

*Guide IV — Un milieu propice à l'apprentissage*, Ottawa, octobre 1995;

*Guide V — Reconnaissance du mérite*, Ottawa, octobre 1995;

*Guide VI — Sondages auprès des employés*, Ottawa, octobre 1995;

*Guide VII — Normes de service*, Ottawa, octobre 1995;

*Guide VIII — Analyses comparatives et meilleures pratiques*, Ottawa, octobre 1995;

*Guide IX — Communications*, Ottawa, octobre 1995;

*Guide X — Analyse comparative et partage des pratiques exemplaires — Mise à jour du Guide VIII*, Ottawa, mars 1996;

*Guide XI — Gestion efficace des plaintes*, Ottawa, juin 1996;

*Guide XII — Qui est le client? — Document de travail*, Ottawa, juillet 1996;

*Guide XIII — Guide des gestionnaires pour la prestation de services de qualité*, Ottawa, septembre 1996.

Canada, Secrétariat du Conseil du Trésor, *L'évaluation des programmes fédéraux : Répertoire sur l'utilisation des évaluations*, Ottawa, 1991.

Canada, Secrétariat du Conseil du Trésor, *Les normes de service : Un guide pour l'initiative*, Ottawa, février 1995.

Canada, Secrétariat du Conseil du Trésor, *Normes d'évaluation de programmes dans les ministères et organismes fédéraux*, Ottawa, juillet 1989.

- Canada, Secrétariat du Conseil du Trésor, *Pour offrir aux Canadiens et aux Canadiennes des services de qualité : Établissement de normes de service au gouvernement fédéral*, Ottawa, décembre 1994.
- Canada, Secrétariat du Conseil du Trésor, *Pour une fonction d'examen plus efficace — Rapport annuel au Parlement par le Président du Conseil du Trésor*, Ottawa, octobre 1995.
- Canada, Secrétariat du Conseil du Trésor, *Repenser le rôle de l'État : Améliorer la mesure des résultats et de la responsabilisation — Rapport annuel au Parlement par le Président du Conseil du Trésor*, Ottawa, octobre 1996.
- Caron, Daniel, J., «Knowledge Required to Perform the Duties of an Evaluator», *Canadian Journal of Program Evaluation*. Vol. 8, N° 1, avril-mai 1993, p. 59 à 78.
- Chelimsky, Eleanor et William R. Shadish, éd., *Evaluation for the 21<sup>st</sup> Century : A Handbook*, Thousand Oaks : Sage Publications, 1997.
- Chelimsky, Eleanor, éd., *Program Evaluation : Patterns and Directions*, Washington : American Society for Public Administration, 1985.
- Chen, Huey-Tsyh, *Theory-Driven Evaluations*, Thousand Oaks : Sage Publications, 1990.
- Fitzgibbon, C.T. et L.L. Morris, *Evaluator's Kit* (2<sup>e</sup> édition), Thousand Oaks : Sage Publications, 1988.
- Hudson, Joe, *et al.*, éd., *Action Oriented Evaluation in Organizations : Canadian Practices*, Toronto : Wall and Emerson, 1992.
- Krause, Daniel Robert, *Effective Program Evaluation : An Introduction*, Chicago : Nelson-Hall, 1996.
- Leeuw, Frans L., «Performance Auditing and Policy Evaluation : Discussing Similarities and Dissimilarities», *Canadian Journal of Program Evaluation*, Vol. 7, N° 1, avril-mai 1992, p. 53 à 68.
- Love, Arnold J., *Evaluation Methods Sourcebook II*, Ottawa : Société canadienne d'évaluation, 1995.
- Martin, Lawrence L. et Peter M. Kettner, *Measuring the Performance of Human Service Programs*, Thousand Oaks : Sage Publications, 1996.
- Mayne, John, *et al.*, éd., *Advancing Public Policy Evaluation : Learning from International Experiences*, Amsterdam : North-Holland, 1992.
- Mayne, John, «In Defence of Program Evaluation», *Canadian Journal of Program Evaluation*, Vol. 1, N° 2, 1986, p. 97 à 102.

Mayne, John et Eduardo Zapico-Goñi, *Monitoring Performance in the Public Sector : Future Directions from International Experience*, New Brunswick (NJ) : Transaction Publishers, 1996.

Paquet, Gilles et Robert Shepherd, *The Program Review Process : A Deconstruction*, Ottawa : Faculté d'administration de l'Université d'Ottawa, 1996.

Patton, M.Q., *Creative Evaluation*, (2<sup>e</sup> édition), Thousand Oaks : Sage Publications, 1986.

Patton, M.Q., *Practical Evaluation*, Thousand Oaks : Sage Publications, 1982.

Patton, M.Q., *Utilization Focused Evaluation*, (2<sup>e</sup> édition), Thousand Oaks : Sage Publications, 1986.

Perret, Bernard, «Le contexte français de l'évaluation : Approche comparative», *Canadian Journal of Program Evaluation*, Vol. 9, N<sup>o</sup> 2, octobre-novembre 1994, p. 93 à 114.

Posavac, Emil J. et Raymond G. Carey, *Program Evaluation : Methods and Case Studies*, (5<sup>e</sup> édition), Upper Saddle River (NJ) : Prentice-Hall, 1997.

Rossi, P.H. et H.E. Freeman, *Evaluation : A Systematic Approach* (2<sup>e</sup> édition), Thousand Oaks : Sage Publications, 1989.

Rush, Brian et Alan Ogborne, «Program Logic Models : Expanding their Role and Structure for Program Planning and Evaluation», *Canadian Journal of Program Evaluation*. Vol. 6, N<sup>o</sup> 2, octobre-novembre 1991, p. 95 à 106.

Rutman, L. et John Mayne, «Institutionalization of Program Evaluation in Canada : The Federal Level», in M.Q. Patton, éd., *Culture and Evaluation*, Vol. 25 of *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1985.

Ryan, Allan G. et Caroline Krentz, «All Pulling Together : Working Toward a Successful Evaluation», *Canadian Journal of Program Evaluation*, Vol. 9, N<sup>o</sup> 2, octobre-novembre 1994, p. 131 à 150.

Shadish, William R., *et al.*, *Foundations of Program Evaluation : Theories of Practice*, Thousand Oaks : Sage Publications, 1991.

Shea, Michael P. et Shelagh M.J. Towson, «Extent of Evaluation Activity and Evaluation Utilization of CES Members», *Canadian Journal of Program Evaluation*, Vol. 8, N<sup>o</sup> 1, avril-mai 1993, p. 79 à 88.

Société canadienne d'évaluation, Comité de normalisation, «Standards for Program Evaluation in Canada : A Discussion Paper», *Canadian Journal of Program Evaluation*, Vol. 7, N<sup>o</sup> 1, avril-mai 1992, p. 157 à 170.



Tellier, Luc-Normand, *Méthodes d'évaluation des projets publics*, Sainte-Foy : Presses de l'Université du Québec, 1994, 1995.

Thurston, W.E., «Decision-Making Theory and the Evaluator», *Canadian Journal of Program Evaluation*, Vol. 5, N° 2, octobre-novembre 1990, p. 29 à 46.

Wye, Christopher G. et Richard C. Sonnichsen, éd., *Evaluation in the Federal Government : Changes, Trends and Opportunities*, San Francisco : Jossey-Bass, 1992.

Zanakis, S.H., *et al.*, «A Review of Program Evaluation and Fund Allocation Methods within the Service and Government», *Socio-economic Planning Sciences*, Vol. 29, N° 1, mars 1995, p. 59 à 79.

Zúñiga, Ricardo, *L'évaluation dans l'action : choix de buts et choix de procédures*, Montréal : Librairie de l'Université de Montréal, 1992.

## Chapitre 2

# STRATÉGIES D'ÉVALUATION

Le présent chapitre commence par une étude des types de conclusions qu'il est possible de tirer d'une évaluation des résultats d'un programme. Il traite des divers obstacles qui sapent typiquement la validité des conclusions de l'évaluation puis passe à un cadre conceptuel pour l'élaboration des stratégies d'évaluation, avant de se terminer par une analyse de la nécessité d'avoir recours à des stratégies de mesure multiples pour produire des conclusions crédibles.

### 2.1 Inférence causale en évaluation

L'évaluation est censée déterminer les résultats obtenus ou «causés» par un programme. Dans cette section, nous essayons de préciser la signification des déclarations sur les causes des résultats d'un programme; la section suivante est une analyse des problèmes d'établissement d'inférences causales.

Commençons par étudier les genres de résultats qu'un programme peut «causer». Dans le plus simple des cas, le programme produit des changements positifs. Cette interprétation suppose toutefois qu'aucun changement positif n'aurait été constaté en l'absence du programme, ce qui n'est pas nécessairement le cas, car la situation aurait pu s'améliorer ou se détériorer quand même. De même, un programme peut maintenir le statu quo en empêchant la situation de se détériorer, et cela peut être son seul effet positif, de sorte qu'il est essentiel de déterminer son effet *incrémentiel*.

Il s'ensuit que, pour bien comprendre les résultats *causés* par un programme, nous devons savoir ce qui serait arrivé sans l'exécution d'un programme. Cette notion est la clé des inférences causales. Autrement dit, si l'on conclut qu'un programme a produit ou causé un certain résultat, cela signifie que, s'il n'avait pas existé, le résultat ne se serait pas concrétisé. Pourtant, cette interprétation de la *causalité* s'applique plus logiquement à certains programmes qu'à d'autres. Elle vaut particulièrement pour les programmes pouvant être considérés comme des interventions gouvernementales pour modifier le comportement de particuliers ou d'entreprises par l'octroi de subventions, la prestation de services ou l'application de règlements. Dans ces cas-là, il est logique et habituellement possible d'arriver à une estimation de ce qui se serait produit si le programme n'avait pas existé.

Il existe toutefois d'autres programmes (dans les secteurs des services médicaux, du contrôle de la circulation aérienne et de la défense, par exemple) qu'il faut considérer logiquement comme partie intégrante du cadre à l'intérieur duquel notre société et notre économie fonctionnent. Ils tendent à exister dans des contextes où l'État assume le rôle d'intervenant principal. En outre, ils sont habituellement universels, ce qui signifie, dans le langage des économistes, que leurs résultats sont des «biens publics». Leur évaluation pose des difficultés parce qu'ils ne se prêtent pas à un modèle d'évaluation dans lequel on les ramène à des interventions précises. En outre, ce sont des programmes permanents, dont l'envergure est habituellement trop grande pour qu'on puisse leur appliquer des méthodes d'évaluation classiques. Certains programmes peuvent faire exception à la règle, mais il reste qu'il faudrait soulever des questions sur la portée de l'évaluation dans le cadre de l'étude préparatoire, à l'intention du client.

Un des derniers aspects de la causalité présente une importance critique dans les cas où les résultats de l'évaluation doivent influencer sur la prise de décisions. On ne peut généraliser à partir des résultats de programmes que l'évaluation a déterminés à moins que le programme lui-même ne puisse être reproduit. Si le programme ne peut exister qu'à un moment, à un endroit ou dans des conditions données, il devient très difficile d'établir des inférences crédibles sur ce qui se produirait dans l'éventualité où un programme analogue serait mis en œuvre ailleurs dans d'autres circonstances.

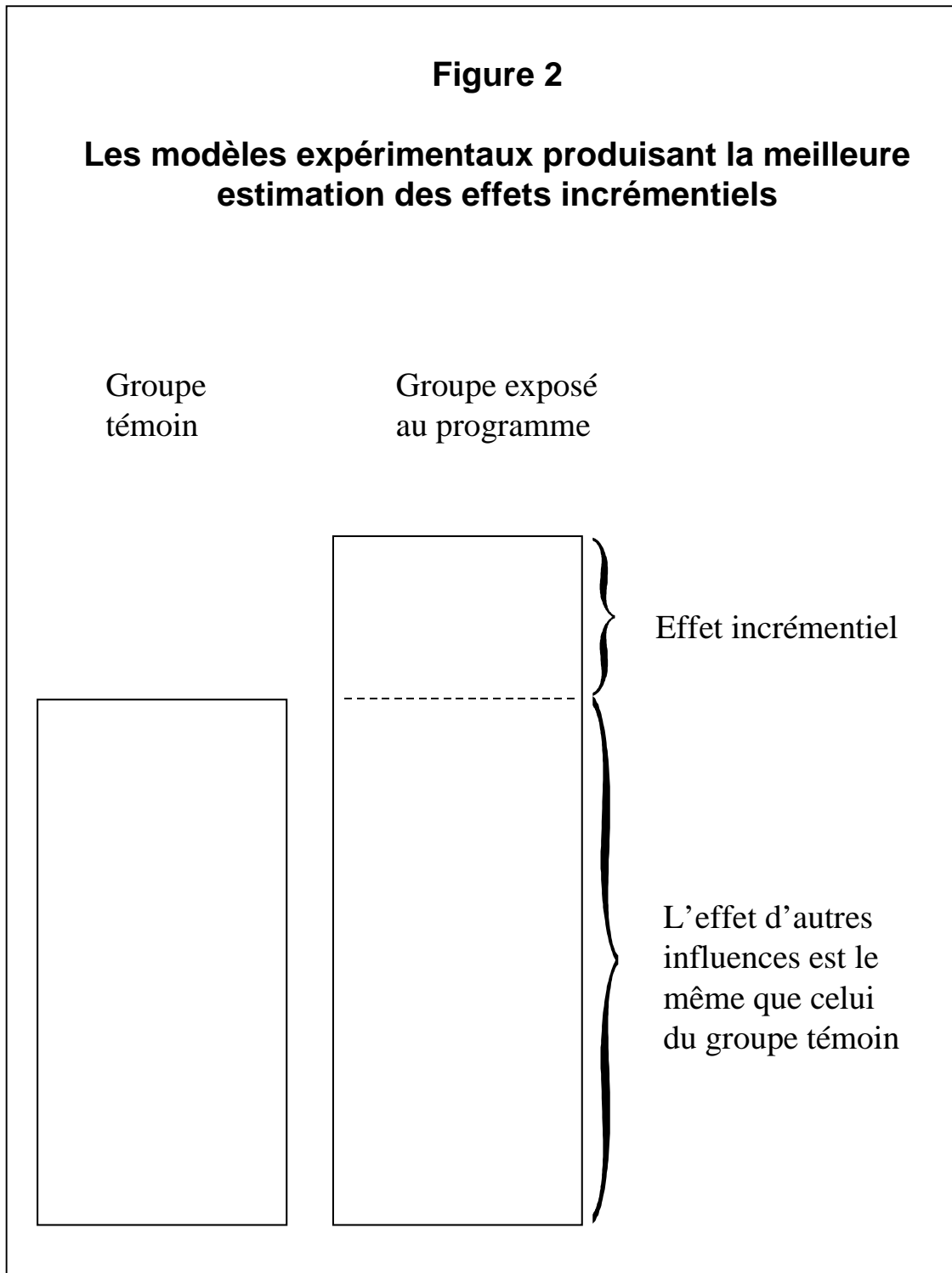
## 2.2 Inférences causales

Conceptuellement, la façon d'établir une inférence causale semble évidente : il suffit de comparer deux situations absolument identiques, exception faite de leur exposition au programme. Toute différence entre deux situations peut être attribuée au programme. Ce principe fondamental est illustré à la figure 2. On choisit deux groupes identiques de sujets (des personnes, des entreprises et des écoles), un seul des deux (le groupe expérimental ou traité) étant exposé au programme, l'autre (le groupe témoin) étant soumis à toutes les mêmes influences extérieures que le groupe expérimental, hormis le programme. Les résultats postérieurs à l'exécution du programme sont mesurés de la même façon pour les deux groupes. À ce moment-là, on peut attribuer au programme n'importe quelle différence entre les résultats des deux groupes, puisque ceux-ci étaient au départ identiques et qu'ils ont été exposés aux mêmes influences extérieures.

Malheureusement, dans la pratique, il n'existe pas de modèle idéal susceptible d'être appliqué parfaitement, puisqu'on ne peut jamais pleinement obtenir l'équivalence absolue du groupe expérimental et du groupe témoin. Des groupes différents sont constitués de sujets différents et diffèrent donc à certains égards, même si les mesures moyennes d'une variable donnée sont identiques. En outre, à supposer que le même groupe serve à la fois de groupe expérimental et de groupe témoin, les observations antérieures et postérieures au programme sont faites à des moments différents, de sorte que d'autres facteurs peuvent influencer sur les résultats observés après l'exécution du programme.

L'impossibilité d'atteindre à l'équivalence absolue sape la validité de l'inférence causale, de sorte qu'il est plus difficile pour les décideurs de déterminer le rendement antérieur du programme et de s'en inspirer dans leurs décisions à l'égard des programmes à venir. C'est d'autant plus complexe que les programmes gouvernementaux ne sont qu'un facteur parmi d'autres qui influent sur les résultats attendus. La rigueur de l'évaluation — et, par conséquent, son utilité pour le processus, décisionnel — est fonction de sa conformité au modèle idéal présenté auparavant.

Il s'ensuit que la capacité d'une inférence selon laquelle le programme a causé un certain résultat dépend, dans la pratique, de la mesure dans laquelle l'évaluation permet de rejeter comme d'autres explications plausibles, souvent appelées «obstacles à la validité de l'inférence causale». D'habitude, l'évaluation ne permet pas d'établir de façon concluante des rapports de cause à effet, mais elle permet de réduire l'incertitude à cet égard tout en produisant des éléments assez concluants pour qu'on puisse réfuter les autres hypothèses. Par exemple, l'évaluation pourrait produire des preuves que le programme est l'explication la plus probable du résultat observé, alors qu'à peu près rien d'autre ne prouve que les autres explications sont valables. Elle pourrait aussi permettre à l'évaluateur de distinguer et de quantifier les effets des autres facteurs en jeu ou des autres explications possibles. **Bref, faire des inférences causales sur les résultats signifie qu'on rejette ou infirme les autres explications plausibles.**



Revenons à notre exemple d'un programme d'aide à l'industrie en vue de créer des emplois. Si nous constatons la création d'un certain nombre d'emplois par les entreprises qui touchent une subvention, nous serions portés à conclure que c'est un

effet attribuable au programme et que si celui-ci n'avait pas existé, les nouveaux emplois n'auraient pas été créés. Néanmoins, avant de pouvoir tirer cette conclusion, il nous faut examiner un certain nombre d'autres explications plausibles. Il se pourrait, par exemple, que les nouveaux emplois aient été créés par suite d'une reprise économique dans le secteur d'activité en question. De même, on pourrait aussi alléguer que les entreprises qui ont créé les emplois avaient l'intention de le faire de toute façon, et que les subventions étaient à toutes fins utiles des paiements de transfert inespérés. Afin de préciser l'effet incrémentiel d'un programme sur la création d'emplois, il faudrait avoir rejeté toutes ces autres explications, voire d'autres encore, ou bien tenir compte de leur influence.

L'élimination des autres explications (celles qui font obstacle à la validité de l'inférence causale posée comme hypothèse) ou l'estimation de leur importance relative est le principal objet d'une évaluation ayant pour but d'établir les résultats d'un programme. C'est une démarche fondée sur une combinaison d'hypothèses, d'éléments logiques et d'analyses empiriques; dans ce manuel, nous appelons chacune de ces approches une *stratégie d'évaluation*.

Revenons encore à l'exemple du programme d'aide à l'industrie : il serait possible de réfuter la conclusion que la création d'emplois résulte d'une reprise économique générale en prouvant qu'il n'y a pas eu de reprise dans la région où l'entreprise est établie (ou dans son secteur de l'économie). Pour ce faire, on étudierait des entreprises du même genre qui n'ont pas reçu de subvention. Si l'on devait constater que des emplois ont été créés uniquement dans celles qui ont touché une subvention, l'explication d'une reprise économique ne serait plus plausible. D'un autre côté, on pourrait remarquer qu'il s'est créé *plus* de nouveaux emplois dans les entreprises qui ont obtenu une subvention que dans les autres, auquel cas il serait toujours possible de rejeter l'explication d'une reprise en attribuant au programme la *différence* entre le nombre d'emplois créés dans les deux groupes d'entreprises (à condition, bien entendu, que les deux groupes se ressemblent suffisamment). Il convient de souligner que cette constatation modifie la conclusion initiale — à savoir que *tous* les nouveaux emplois sont attribuables au programme — compte tenu de l'effet d'une reprise économique. De plus, malgré ses limitations, ce modèle de comparaison permet d'éliminer bon nombre d'explications, y compris celle que les entreprises auraient créé les emplois en question de toute façon. Dans cet exemple, si ces deux autres explications sont les seules qu'on juge vraisemblables, la conclusion que le nombre accru de nouveaux emplois est attribuable au programme deviendrait assez plausible, d'après les éléments de preuve présentés. Toutefois, comme nous le verrons au chapitre suivant, il y a de plus fortes chances que les deux groupes d'entreprises n'aient pas été tout à fait semblables, de sorte que d'autres obstacles sapent la validité des conclusions. En pareil cas, il faut élaborer d'autres stratégies d'évaluation pour éliminer ces obstacles.

Jusqu'ici, nous avons tenté de déterminer dans quelle mesure un programme produit un résultat observé. Il reste un autre facteur qui vient compliquer l'équation : même si le programme est indispensable pour que le résultat se produise, il n'est pas nécessairement suffisant. Autrement dit, le résultat peut aussi être attribuable à

d'autres facteurs, en l'absence desquels il n'est pas atteint. Sans le programme, il n'y a pas de résultat, mais cela ne signifie pas nécessairement que son existence assurera le résultat désiré. Tout ce qu'on peut déduire, c'est que le résultat se produira si le programme est mis en œuvre *et* que les autres facteurs favorables sont réunis.

L'intérêt de ces autres facteurs s'explique du fait que, lorsqu'on a abouti à une conclusion au sujet de l'effet d'un programme existant, on veut normalement la généraliser en l'appliquant à d'autres lieux, à d'autres moments ou à d'autres situations. Cette possibilité de généraliser, appelée la *validité externe* de l'évaluation, se limite à affirmer que, dans des conditions identiques, la mise en œuvre du programme ailleurs entraînerait le même résultat. Bien sûr, ni les conditions, ni le programme ne peuvent être parfaitement reproduits, de sorte que les inférences de ce genre sont souvent chancelantes au point que, pour les rendre crédibles, il faut poser de nouvelles hypothèses, trouver d'autres arguments logiques ou réaliser d'autres analyses empiriques. Il peut alors être utile d'avoir recours à des stratégies d'évaluation multiples.

Revenons une fois de plus à l'exemple du programme de subventions à l'industrie. Qu'arrivera-t-il si nous devons établir que le programme existant a effectivement permis de créer un certain nombre d'emplois, grâce à certaines compétences en marketing et à d'autres facteurs? Ce résultat peut être utile du point de vue de la responsabilisation, mais les questions posées au sujet de l'élaboration de nouveaux programmes devraient alors normalement porter sur l'opportunité de poursuivre le programme, de lui donner de l'expansion ou d'en réduire l'ampleur. La validité externe de la conclusion selon laquelle la poursuite ou l'expansion du programme entraînerait la création de nouveaux emplois pourrait être sujette à caution si l'échantillon des entreprises étudiées n'était pas représentatif de toutes celles auxquelles le programme s'appliquerait, ou si les conditions qui ont contribué au succès du programme dans le passé étaient peu susceptibles de se reproduire. Il se pourrait que les autres entreprises n'aient pas les aptitudes en marketing nécessaires, de sorte que le programme élargi n'aurait pas un effet comparable sur elles. Bref, c'est compte tenu de la question à l'étude et du genre de décisions à prendre que l'évaluateur pourra cerner d'autres facteurs explicatifs et explorer leurs liens avec le programme.

Il existe diverses stratégies pour qui veut minimiser l'effet des obstacles à la validité externe, tout comme à la validité interne, d'ailleurs. Malheureusement, elles ne sont pas toujours compatibles, de sorte qu'il faut parfois opter pour une solution de compromis. Quand l'évaluateur doit formuler des conclusions crédibles sur lesquelles la direction peut se fonder utilement, il est clair que, malgré l'importance indéniable de la validité interne, la validité externe de l'évaluation ne saurait être négligée. L'évaluateur devrait toujours être conscient du genre de décisions à prendre et, partant, du genre de conclusions qu'il doit présenter. Il doit donc bien comprendre les principaux obstacles à la validité externe, si des points ne sont pas traités, ainsi qu'à la crédibilité et à l'utilité de ces conclusions pour les décideurs.

## Résumé

Les difficultés d'établissement d'inférences causales quant aux programmes et à leurs résultats sont l'un des principaux thèmes du manuel. L'autre thème principal est celui de la mesure des résultats. Avant de pouvoir tirer des conclusions sur les effets d'un programme, l'évaluateur doit être conscient des autres facteurs ou des autres circonstances susceptibles d'expliquer les résultats observés, puis présenter des arguments pour réfuter ces explications. S'il fait des généralisations à partir de ses conclusions, il devrait surveiller de près les obstacles à la validité externe de son évaluation. Les méthodes utilisées pour déterminer les résultats d'un programme sont bonnes dans la mesure où elles permettent de produire les meilleurs arguments possibles, compte tenu des ressources et du temps disponibles.

## Références : Inférence causale

Campbell, D.T. et J.C. Stanley, *Experimental and Quasi-experimental Designs for Research*, Chicago : Rand-McNally, 1963.

Cook, T.D. et D.T. Campbell, *Quasi-experimentation : Design and Analysis Issues for Field Settings*, Chicago : Rand-McNally, 1979.

Cook, T.D. et C.S. Reichardt, éd., *Qualitative and Quantitative Methods in Evaluation Research*, Thousand Oaks : Sage Publications, 1979.

Heise, D.R., *Causal Analysis*, New York : Wiley, 1985.

Kenny, D.A., *Correlation and Causality*, Toronto : John Wiley and Sons, 1979.

Suchman, E.A., *Evaluative Research : Principles and Practice in Public Service and Social Action Programs*, New York : Russell Sage, 1967.

Williams, D.D., éd., *Naturalistic Evaluation*, Vol. 30 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1986.

## Notes :



## 2.3 Stratégies d'évaluation

Il faut tenir compte de deux types de facteurs afin d'élaborer des méthodes de détermination des résultats d'un programme : les facteurs qui sont liés à la *recherche* (autrement dit à la qualité des éléments de preuve réunis) et les facteurs qui découlent du *contexte décisionnel* dans lequel l'évaluation a lieu. Les deux facteurs sont importants, mais, quoi qu'il en soit, il faut ordinairement concilier la rigueur scientifique de l'évaluation et sa pertinence pour les décideurs.

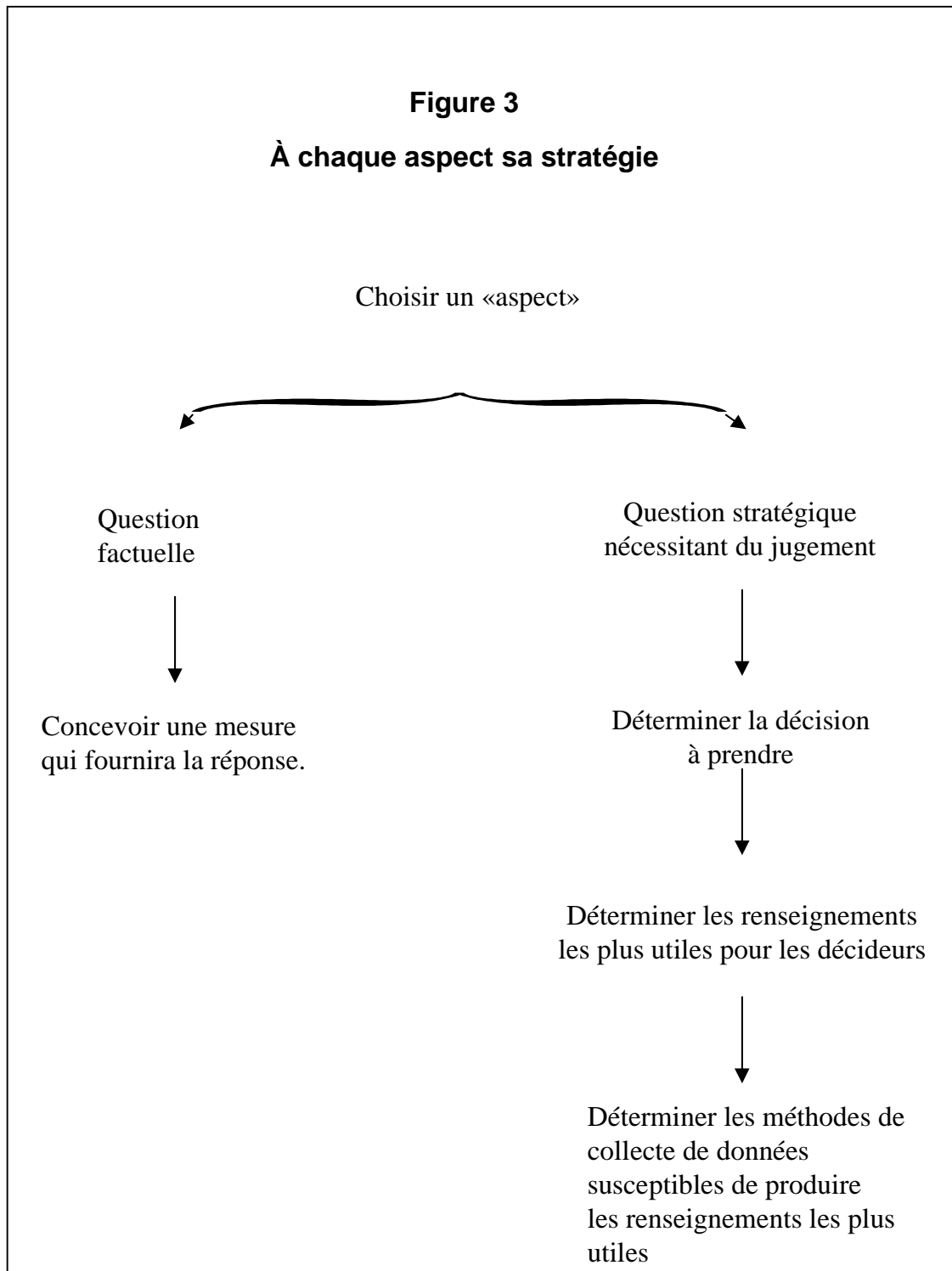
Il existe plusieurs façons de recueillir les données sur lesquelles on se fonde pour déterminer les résultats d'un programme. Dans ce chapitre, nous étudions les principales *stratégies d'évaluation*, qui comprendront toutes un *modèle d'évaluation* (chapitre 3), une *méthode de collecte des données* (chapitre 4) et une *méthode analytique* (chapitre 5).

Dans notre exemple du programme d'aide à l'industrie, on pourrait décider de déterminer si les emplois créés sont attribuables au programme en menant une enquête auprès des entreprises participantes pour leur demander ce qui serait arrivé s'il n'y avait pas eu de subvention gouvernementale. Une autre stratégie pourrait consister à faire un sondage pour déterminer le nombre d'emplois créés dans des entreprises analogues, les unes ayant reçu une subvention et les autres pas, puis à comparer les résultats afin de mesurer les importantes différences statistiques. Une troisième stratégie pourrait faire appel à des études de cas approfondies sur des entreprises ayant bénéficié d'une subvention pour déterminer si elles auraient vraisemblablement créé les emplois en question de toute façon. Chacune de ces stratégies porte sur la même question et fournit des preuves de nature et de qualité différentes; aucune ne fournit normalement de preuve incontestable des résultats du programme. C'est pourquoi il est donc souvent approprié d'avoir recours à plusieurs stratégies. Par exemple, on peut vouloir aussi déterminer les effets du programme à d'autres égards, celui de la concurrence déloyale que les subventions auraient pu créer. Cela pourrait se faire en partie au moyen d'une des stratégies susmentionnées, et en partie aussi grâce à une stratégie différente. La stratégie globale pour laquelle l'évaluateur opte est le plus souvent une combinaison de stratégies différentes conçue pour trancher une série de questions précises. À la section 2.4.3, nous verrons comment on élabore de telles stratégies ou des démarches d'évaluation multiples.

La figure 3 illustre les étapes générales de l'élaboration d'une stratégie d'évaluation. Il est utile d'envisager cette démarche comme une série d'étapes que nous décrivons dans l'ordre, bien qu'elle soit beaucoup plus itérative dans la pratique, puisque chaque étape est étroitement liée aux autres.

Pour commencer, l'évaluateur doit choisir un modèle. Le *modèle d'évaluation* s'entend du modèle logique utilisé pour parvenir à des conclusions sur les résultats. Afin de le choisir, l'évaluateur doit déterminer simultanément le genre d'information qu'il veut obtenir et le type d'analyse auquel il va soumettre cette information. Par exemple, si l'évaluation a pour objet de déterminer dans quelle mesure un programme a atteint un objectif donné, l'évaluateur doit choisir un indicateur approprié et opter pour une méthode d'analyse qui lui permettra d'isoler l'effet du programme. Les

modèles d'évaluation servent de base logique pour mesurer des résultats et les attribuer aux programmes.



Une fois le modèle d'évaluation choisi, l'évaluateur passe au choix des méthodes et des techniques nécessaires pour l'appliquer. Le genre d'information nécessaire — indicateurs qualitatifs ou indicateurs quantitatifs de la réalisation des objectifs — est déterminé à l'étape de la conception du modèle. L'étape suivante consiste à définir les *données* nécessaires pour produire cette information. Les données sont des faits, c'est-à-dire des choses qu'on peut observer et consigner et leur nature et leur qualité peuvent varier nettement. À cet égard, la tâche de l'évaluateur se complique parce que les données sont plus ou moins accessibles et que leur coût et leur pertinence varient. C'est là qu'intervient la question de la *mesure*, puisqu'il faut décider quelles données sont les plus pertinentes et comment les recueillir. Comme nous le verrons ultérieurement, la mesure est une question méthodologique d'importance cruciale pour l'évaluation.

Après avoir bien défini les données nécessaires, l'évaluateur doit déterminer leurs sources potentielles. S'il lui est impossible d'obtenir des données fiables d'une source secondaire, il doit avoir recours à une méthode de collecte de données primaires (Cook et Campbell, 1970, chapitre 1; Cronbach, 1982, chapitre 4). Cette approche est généralement plus coûteuse que celle de l'utilisation de données secondaires, et on devrait s'en écarter dans la mesure du possible. Lorsqu'on décide de recueillir des données primaires, il faut normalement choisir une méthode de collecte (observations sur le terrain et sondages postaux, par exemple), mettre au point des instruments de mesure (questionnaires, guides d'entrevue, fiches d'enregistrement des observations, etc.) et formuler un plan d'échantillonnage.

Enfin, compte tenu du type d'analyse nécessaire et du genre de données disponibles, l'évaluateur doit choisir des méthodes d'*analyse des données* (analyse coûts-avantages, régression multiple, analyse de la variance, etc.). Ces analyses ont pour objet de traduire les données recueillies pour produire l'information nécessaire pour l'évaluation.

**Notes :**

## 2.4 Évaluations crédibles

Avant d'analyser les éléments précis d'une stratégie d'évaluation de façon plus détaillée, nous devrions examiner les éléments clés dont on doit tenir compte pour assurer la crédibilité de l'évaluation elle-même. Ces éléments clés sont résumés au tableau 2.

<b>Tableau 2</b>	
<b>Éléments nécessaires à la crédibilité des évaluations</b>	
<b>A.</b>	<b>Critères de recherche</b>
	<ul style="list-style-type: none"><li>• questions de mesure<ul style="list-style-type: none"><li>– fiabilité</li><li>– validité de la mesure</li><li>– profondeur et portée</li></ul></li><li>• questions d'attribution<ul style="list-style-type: none"><li>– validité des inférences causales</li></ul></li></ul>
<b>B.</b>	<b>Critères du contexte décisionnel</b>
	<ul style="list-style-type: none"><li>• possibilité de formuler des conclusions crédibles<ul style="list-style-type: none"><li>– objectivité</li><li>– pertinence pour le contexte décisionnel</li><li>– pertinence du niveau et du genre de données</li><li>– exhaustivité</li></ul></li><li>• questions pratiques<ul style="list-style-type: none"><li>– faisabilité</li><li>– coût abordable</li><li>– éthique</li></ul></li></ul>

### 2.4.1 Critères de recherche

#### a) Questions de mesure

Bien des effets des programmes sont fondamentalement difficiles à mesurer. Voici quelques exemples :

- amélioration du bien-être des personnes âgées, grâce à des programmes leur permettant de continuer à vivre seules chez elles;

- amélioration de la sécurité nationale grâce à la mise au point d'un important système d'armes;
- amélioration des stimulants à la R-D industrielle grâce à des modifications du régime fiscal.

Comme bien d'autres, ces effets exigent à la fois des méthodes de mesure perfectionnées et une connaissance approfondie de domaines spécialisés de la politique publique.

Trois des aspects de la mesure nécessitent une attention particulière : la fiabilité, la validité de la mesure, ainsi que la profondeur et la portée.

### **Fiabilité**

Une mesure est dite fiable si elle donne les mêmes résultats lorsqu'elle est appliquée de façon répétée dans une situation donnée. Par exemple, un test d'intelligence serait considéré comme fiable dans la mesure où il donnerait un résultat identique s'il était administré deux fois à la même personne (dont l'intelligence n'aurait pas changé). Dans le contexte d'un programme, la fiabilité peut correspondre à la stabilité de la mesure dans le temps, ou à son uniformité d'un endroit à l'autre.

Le manque de fiabilité peut être attribuable à plusieurs facteurs. Par exemple, il peut résulter d'une mauvaise méthode de collecte des données : si l'enquêteur ne lit pas attentivement les instructions du guide d'entrevue, il risque d'obtenir des résultats légèrement différents de ceux des enquêteurs qui les ont lues. Le manque de fiabilité peut aussi être imputable à l'instrument de mesure lui-même, ou au plan d'échantillonnage. Si la procédure d'échantillonnage n'est pas bien suivie, l'échantillon risque de ne pas être représentatif de la population visée et, par conséquent, les résultats qu'il génère peuvent n'être pas fiables.

### **Validité de la mesure**

Une mesure est valide dans la mesure où elle représente fidèlement ce qu'elle est censée représenter. Les mesures (indicateurs) valides ne présentent pas d'erreurs systématiques et saisissent les données voulues. Les données signifient-elles ce que nous croyons qu'elles signifient? La technique employée mesure-t-elle ce qu'elle est censée mesurer? Ces questions ont une importance critique pour l'évaluation des programmes.

Les problèmes de validité d'une mesure peuvent être conceptuels ou techniques. À moins d'y avoir bien réfléchi, il est rare qu'on sache exactement quelles données correspondent le mieux aux résultats à mesurer. La décision est trop souvent fondée uniquement sur des données faciles à obtenir, mais qui génèrent des mesures moins probantes qu'on n'aurait pu le souhaiter. En outre, des erreurs techniques (de mesure et d'échantillonnage, par exemple) peuvent se produire et fausser les résultats de l'évaluation.

## Profondeur et portée

Les notions de profondeur et de portée sont indissociables de celles de la fiabilité et de la validité de la mesure. Dans certaines situations, l'évaluateur peut souhaiter mesurer certains résultats très précisément et d'autres de façon moins détaillée, mais avec plusieurs instruments différents.

Pour mesurer les avantages d'un programme pour une personne, il faut parfois réaliser des entrevues et des sondages en profondeur. Dans certains cas, il peut arriver aussi qu'on doive avoir recours à différents indicateurs reflétant tous des points de vue distincts sur les conséquences envisagées. Par exemple, lorsqu'on évalue l'effet d'une subvention pour une entreprise, il peut être nécessaire d'analyser ses ventes, l'évolution de son effectif, celle de la qualité de ses emplois, l'effet de l'achat de nouvelles machines sur sa compétitivité, et ainsi de suite.

Par ailleurs, la population cible d'un programme peut être importante et hétérogène, auquel cas il est important que l'évaluation porte de façon relativement peu détaillée sur tous ses segments. Ainsi, pour évaluer convenablement les conséquences pour les entreprises d'un programme d'aide à un secteur d'activité donné, il faudrait prendre soin d'assurer une représentation suffisante de tous les types d'entreprises visées (grandes et petites, de différents secteurs et de régions différentes).

La profondeur et la portée de la mesure posent un problème épineux. Comme le temps et les ressources sont limités, l'évaluateur doit inévitablement négliger l'une au profit de l'autre. S'il privilégie la portée, son évaluation peut gagner en pertinence et avoir un champ d'application plus vaste, mais elle perd alors en profondeur, et les mesures individuelles sont alors moins valides et moins fiables.

### b) Questions d'attribution

Le programme n'est souvent qu'un des nombreux facteurs influant sur le résultat constaté. En fait, il peut être assez difficile de déterminer dans quelle proportion les résultats sont vraiment attribuables au programme plutôt qu'à d'autres facteurs. C'est peut-être l'aspect le plus difficile d'une étude d'évaluation.

La clé de l'attribution des résultats est donc une bonne comparaison. En laboratoire, il est possible de le faire grâce à des groupes témoins rigoureusement contrôlés. Par contre, dans le cas des programmes du gouvernement fédéral, les comparaisons qui sont généralement possibles sont moins rigoureuses, et de nombreux obstacles risquent de saper la *validité interne* et la *validité externe*.

Les **obstacles à la validité interne** les plus courants sont les suivants :

- **événements historiques** — événements externes influant sur les participants au programme autrement que sur les membres des groupes témoins;

- **maturation** — changements des résultats découlant du temps écoulé plutôt qu'attribuables au programme lui-même (p. ex., le vieillissement des participants d'un groupe comparativement à ceux d'un autre à une étape différente);
- **attrition** — abandon du programme par des répondants (ce facteur pourrait nuire à la comparabilité des groupes expérimental et témoin);
- **biais de la sélection** — propension initialement inégale des groupes expérimental et témoin à réagir au programme;
- **facteurs de régression** — pseudo-changements des résultats découlant de la rétention de personnes pour un programme en raison des résultats extrêmes obtenus (à la longue, tout groupe extrême a tendance à se rapprocher de la moyenne, qu'il ait bénéficié du programme ou pas);
- **diffusion ou imitation du traitement** — obtention par les répondants d'un groupe de l'information destinée à un autre groupe;
- **essai** — différences observées entre les groupes expérimental et témoin pouvant être imputables à une meilleure connaissance d'un instrument de mesure pour les membres du premier groupe;
- **instruments** — conséquence du changement de l'instrument utilisé selon le groupe (p. ex., lorsqu'on a recours à différents enquêteurs).

Il existe aussi de nombreux **obstacles à la validité externe**, autrement dit empêchant l'évaluateur de généraliser ses constatations pour les appliquer dans d'autres contextes, à d'autres moments ou à d'autres programmes. Dans le contexte de l'administration fédérale, la validité externe a toujours une grande importance, puisque les constatations de l'évaluation sont censées appuyer la prise de décisions ultérieures.

Il existe trois types d'obstacles à la généralisation des constatations :

- **interaction entre la sélection et le programme** — non-représentativité des effets sur les participants au programme, parce que ceux-ci ont une caractéristique (influant sur les effets) non représentative de l'ensemble de la population;
- **interaction entre le contexte et le programme** — non-représentativité du contexte du programme expérimental ou pilote comparativement à celui dans lequel le programme aurait été exécuté, s'il avait été entièrement mis en œuvre;
- **interaction entre les événements historiques et le programme** — non-représentativité des conditions dans lesquelles le programme s'est déroulé par rapport aux conditions futures.

Lorsqu'on est appelé à choisir des stratégies d'évaluation, il est manifestement très utile d'être conscient des obstacles à leur validité. Une grande partie du jugement qui est nécessaire à la conception d'une évaluation ainsi qu'à la collecte et à l'analyse des données consiste à savoir trouver les moyens de déterminer les effets attribuables au programme. Pour y arriver, il faut établir de bonnes comparaisons, en évitant de donner prise au plus grand nombre d'obstacles possible à la validité.

Lorsque l'évaluation est axée sur les résultats, les modèles diffèrent surtout quant à l'efficacité avec laquelle ils permettent de déterminer les effets attribuables au programme et, le cas échéant, à la facilité de généralisation des conclusions. Les modèles d'évaluation sont présentés au chapitre 3, en ordre décroissant de crédibilité.

### **Références : Stratégies d'évaluation**

Campbell, D.T. et J.C. Stanley, *Experimental and Quasi-experimental Designs for Research*, Chicago : Rand-McNally, 1963.

Cook, T.D. et D.T. Campbell, *Quasi-experimentation : Designs and Analysis Issues for Field Settings*, Chicago : Rand-McNally, 1979.

Kerlinger, F.N., *Behavioural Research : A Conceptual Approach*, New York : Holt, Rinehart and Winston, 1979, chapitre 9.

Mercer, Shawna L. et Vivek Goel, «Program Evaluation in the Absence of Goals : A Comprehensive Approach to the Evaluation of a Population-Based Breast Cancer Screening Program», *Canadian Journal of Program Evaluation*, Vol. 9, N° 1, avril-mai 1994, p. 97 à 112.

Patton, M.Q., *Utilization-focussed Evaluation* (2<sup>e</sup> édition), Thousand Oaks : Sage Publications, 1986.

Rossi, P.H. et H.E. Freeman, *Evaluation : A Systematic Approach* (2<sup>e</sup> édition), Thousand Oaks : Sage Publications, 1989.

Ryan, Brenda et Elizabeth Townsend, «Criteria Mapping», *Canadian Journal of Program Evaluation*, Vol. 4, N° 2, octobre-novembre 1989, p. 47 à 58.

Watson, Kenneth, «Selecting and Ranking Issues in Program Evaluations and Value-for-money Audits», *Canadian Journal of Program Evaluation*, Vol. 5, N° 2, octobre-novembre 1990, p. 15 à 28.

### **Notes :**



### 2.4.2 Critères du contexte décisionnel

Puisque l'évaluation est censée faciliter la prise de décisions, les critères de choix d'une méthode d'évaluation appropriée doivent garantir l'obtention d'une information *utile*. Cela suppose qu'on comprenne le contexte dans lequel les décisions seront prises et où les constatations de l'évaluation seront présentées. Il faut donc tenir compte de facteurs qui s'ajoutent aux aspects techniques des méthodes, bien que celles-ci conservent une importance critique pour la crédibilité des constatations.

L'élaboration d'une démarche d'évaluation des résultats d'un programme peut donc devenir une tâche très délicate, qui tient probablement plus de l'art que de la science, puisqu'il faut tenir compte à la fois des avantages et des inconvénients des stratégies envisagées pour recueillir des données et du contexte dans lequel l'évaluation se déroule. La conciliation de ces deux éléments doit en outre se faire en fonction des contraintes imposées par les ressources et le temps limité dont l'évaluateur dispose. Bref, c'est une tâche qui exige de toute évidence l'expérience de la recherche, et de la gestion.

Lorsqu'on examine les démarches d'évaluation possibles à l'étape préparatoire de la planification, il faudrait constamment se poser la question suivante : la méthode ou solution recommandée fournira-t-elle des données suffisantes sur les questions visées, dans les délais fixés et sans dépasser le budget? Le tableau 2 présente deux éléments dont il faut se rappeler dans le contexte décisionnel : le degré auquel on peut s'attendre que la méthode aboutisse à des conditions crédibles et celui auquel elle peut être appliquée. Nous allons maintenant décrire chacun de ces éléments généraux en traitant aussi des questions connexes qui sont décrites ci-dessous. Il convient de souligner que ces éléments s'appliquent à tous les aspects de l'évaluation, pas seulement à ceux qui sont liés aux résultats du programme.

#### a) **Formulation de conclusions crédibles (recommandations judicieuses fondées sur une analyse précise)**

La démarche d'évaluation devrait tenir compte de la possibilité de formuler des conclusions crédibles.

On recueille des données afin de formuler des conclusions objectives et crédibles basées sur elles, avec assez de preuves à l'appui pour qu'on y ajoute foi. Il peut être difficile d'aboutir à de telles conclusions, et l'évaluateur devrait en tenir compte lorsqu'il élabore sa stratégie. En outre, la crédibilité des conclusions est en partie fonction de leur formulation, autrement dit de leur présentation.

Les données recueillies et les conclusions formulées devraient être objectives, et toutes les hypothèses devraient être clairement précisées.

L'objectivité des évaluations est extrêmement importante. En effet, elles sont souvent contestées par quelqu'un, soit un gestionnaire de programme, un client, un membre de la haute direction, un représentant d'un organisme central ou un ministre. L'objectivité signifie que les données et les conclusions peuvent être vérifiées et confirmées par d'autres personnes que les auteurs de l'évaluation. Autrement dit, les conclusions doivent découler de l'information recueillie. L'information et les données d'évaluation devraient donc être réunies, analysées et présentées de telle façon que d'autres personnes qui feraient la même évaluation en se fondant sur les mêmes hypothèses de base aboutiraient à des conclusions analogues. C'est beaucoup plus difficile à faire lorsqu'on opte pour certaines stratégies d'évaluation que pour d'autres, notamment si la stratégie utilisée repose largement sur le jugement professionnel de l'évaluateur. En particulier, on devrait toujours préciser clairement au lecteur les éléments sur lesquels les conclusions sont fondées (l'information et les données recueillies ainsi que les hypothèses posées). Si les conclusions sont ambiguës, il est particulièrement important que les hypothèses sous-jacentes soient clairement énoncées. En effet, lorsqu'elles ne sont pas bien précisées, il arrive souvent que les conclusions soient mal formulées.

Les conclusions doivent être pertinentes, c'est-à-dire compatibles avec le contexte décisionnel et elles doivent absolument porter sur les questions étudiées.

Au cours d'une étude, les chercheurs perdent parfois de vue les questions sur lesquelles l'examen doit porter; il devient alors difficile pour le lecteur (le client de l'évaluation) de comprendre le lien entre les conclusions et les questions à évaluer cernées au départ. Ce phénomène peut être dû à plusieurs facteurs. Il se peut par exemple que la stratégie d'évaluation n'ait pas été suffisamment bien conçue, de sorte qu'il est difficile d'obtenir de l'information valide sur certaines questions et de tirer certaines conclusions. Par ailleurs, il est possible aussi que les intérêts de l'évaluateur l'emportent, auquel cas les questions qui intéressent la haute direction ne reçoivent pas toute l'attention voulue. Enfin, d'autres questions peuvent se poser pendant qu'on étudie le programme et son contexte. Cela ne devrait toutefois pas présenter de difficulté, pourvu que les questions initiales soient bel et bien étudiées et qu'on précise clairement les questions supplémentaires et les conclusions correspondantes.

La précision des conclusions est largement fonction de la qualité et de la nature de l'information recueillie, lesquelles devraient être choisies compte tenu des facteurs contextuels.

On constate souvent deux types de difficultés dans les évaluations. Il est souvent impossible d'arriver à des conclusions définitives, et l'information et les données recueillies grâce aux stratégies utilisables ne sont pas complètes.

Dans le premier cas, il est fréquent qu'on n'arrive pas à prouver catégoriquement le rapport de causalité entre un programme et un résultat observé, en raison surtout de l'impossibilité de surmonter les problèmes de mesure et d'attribution dont nous avons déjà fait état. En général, il est peu probable qu'une stratégie d'évaluation produise à elle seule suffisamment d'informations pour donner une réponse sans équivoque aux questions posées.

Cela nous amène directement au second type de difficultés : il y a normalement plusieurs stratégies d'évaluation envisageables, chacune produisant de l'information et des données de qualité et de nature différentes. **Il s'ensuit donc qu'il faudrait choisir la stratégie en se fondant sur les facteurs contextuels liés aux décisions à prendre au sujet du programme, et pas seulement sur des questions de recherche prédéfinies.** C'est sensiblement la même chose qu'en droit, où le genre d'éléments de preuve à produire est fonction de la gravité et du type de crime. Ainsi, dans bien des poursuites au civil, il suffit de prouver l'existence de motifs raisonnables, alors que la culpabilité d'un criminel doit être prouvée «au-delà de tout doute raisonnable» (Smith, 1981). Les facteurs contextuels dont l'évaluateur devrait tenir compte sont le degré d'incertitude sur le programme et sur ses résultats, l'importance de ses effets, son coût et la probabilité que les conclusions soient contestées. Il devrait être capable de prévoir quelles contestations d'envergure ses conclusions susciteront éventuellement, et être prêt à les réfuter.

Le choix de l'information à recueillir et, partant, de la méthode d'évaluation à utiliser est l'une des tâches les plus difficiles pour l'évaluateur. En principe, c'est le client de l'étude et *non* l'évaluateur qui fera ce choix. La tâche de l'évaluateur consiste à présenter au client les démarches d'évaluation susceptibles de générer les conclusions crédibles qu'on attend de lui, à un coût et dans des délais raisonnables. Pour choisir la démarche, le client devrait avoir une bonne compréhension de l'information qui sera produite et être par conséquent en mesure de juger si l'évaluation est suffisamment rigoureuse pour pouvoir s'en inspirer dans ses décisions. Bien entendu, l'évaluateur devrait proposer des démarches d'évaluation qui reflètent le mieux possible le contexte décisionnel afin de faciliter le choix du client.

Les conclusions formulées devraient être fondées sur un examen exhaustif des questions pertinentes.

L'exhaustivité — ou son absence — est un autre facteur qui pose souvent des problèmes aux évaluateurs. (Bien qu'elle soit liée à la pertinence de l'information, elle constitue un point distinct dans le tableau 2, parce qu'on a souvent tendance à produire de l'information et des données objectives et pertinentes sur la plupart des questions à l'étude, mais à en négliger plus ou moins d'autres.) Il s'agit là d'un problème de macromesure. L'évaluateur devrait s'efforcer d'avoir une idée aussi exacte que possible de la question *du point de vue du client*. Cela suppose qu'il étudie toutes les questions d'intérêt qu'il peut, compte tenu du temps et des ressources financières dont il dispose. À cet égard, il ne faut jamais oublier que, pour le gouvernement fédéral, le «client» est en définitive le public canadien. Il est parfois difficile de faire en sorte que la portée de l'évaluation soit suffisante. Pourtant, si l'on

décide de la sacrifier pour analyser de façon plus approfondie certaines des questions envisagées, on risque d'aboutir à des conclusions correctes, mais sans vue d'ensemble. Pour éviter cet écueil, on prend habituellement soin de discuter des questions d'évaluation avec le client et avec d'autres parties ayant des points de vue différents. De cette façon, on a toutes les chances d'arriver à une stratégie d'évaluation d'une portée satisfaisante.

Si l'évaluateur estime que sa tâche consiste à fournir un complément d'information pertinente sur un programme et sur ses résultats (autrement dit de proposer une méthode permettant de réduire l'incertitude au sujet d'un programme) plutôt qu'à produire des preuves concluantes de son efficacité, il aboutira donc vraisemblablement à des conclusions plus utiles. Avec cette approche, il risque de devoir faire des choix difficiles entre la pertinence et la rigueur de son travail, mais il doit choisir des méthodes d'évaluation qui lui permettront de maximiser les chances d'arriver à des conclusions utiles, même avec des réserves.

Enfin, on devrait clairement distinguer les constatations des recommandations de l'évaluation.

L'évaluateur peut être fréquemment appelé à donner des conseils à son client et à lui présenter des recommandations. Il doit alors absolument établir une distinction entre les constatations qui sont tirées de l'information générée par son étude et les recommandations sur le programme qui s'inspirent des conclusions de son évaluation ou de renseignements provenant d'autres sources, par exemple des directives stratégiques. Les conclusions de l'évaluation perdent de leur crédibilité si cette distinction n'est pas maintenue.

Par exemple, les constatations d'une évaluation d'un programme d'économie d'énergie résidentielle peuvent permettre à l'évaluateur de conclure que le programme a eu des répercussions favorables sur l'économie d'énergie. Toutefois, des renseignements obtenus d'autres sources peuvent laisser entendre que d'autres programmes d'économie d'énergie sont plus rentables, auquel cas l'évaluateur est porté à recommander que le programme résidentiel soit abandonné. Dans ce cas-là, il doit clairement préciser que sa recommandation n'est pas fondée sur l'information obtenue dans le contexte de l'évaluation elle-même, mais bien sur d'autres renseignements.

## **b) Questions pratiques**

Lorsqu'il élabore sa méthode d'évaluation, l'évaluateur doit tenir compte d'éléments fondamentaux tels que la praticabilité, la viabilité financière et l'éthique.

Une démarche est jugée *praticable* dans la mesure où elle peut être appliquée efficacement sans conséquences néfastes et dans les délais impartis. La *viabilité financière* s'entend du coût de mise en œuvre de la démarche. Il se peut que le coût d'utilisation de la méthode considérée comme la plus appropriée dans une situation donnée soit exorbitant. Or, il faut toujours préférer la méthode d'évaluation susceptible à la fois de gérer les problèmes de mesure et d'attribution et d'aboutir à des conclusions crédibles, tout en pouvant être appliquée dans les limites des ressources disponibles.

L'*éthique* (principes ou valeurs morales) doit être évaluée dans l'élaboration d'une méthode d'évaluation. Par exemple, il peut être contraire à l'éthique d'exécuter un programme exclusivement pour un sous-groupe d'une population donnée. Ce serait le cas si une évaluation portant sur un programme social devait être fondée sur un échantillon aléatoire de prestataires et privait de services d'autres personnes y ayant pourtant autant droit. Les principes d'éthique dont il faut tenir compte dans le contexte des évaluations de programmes de l'administration fédérale sont précisés dans divers textes législatifs et stratégiques sur la collecte, l'utilisation, la préservation et la diffusion de l'information, dont la *Loi sur l'accès à l'information*, la *Loi sur la protection des renseignements personnels* et la *Loi sur la statistique*, ainsi que la Politique du gouvernement en matière de communications et la Politique sur la gestion des renseignements détenus par le gouvernement du Conseil du Trésor, laquelle porte notamment sur les mesures à prendre pour minimiser la collecte de données inutile et pour assurer l'examen méthodologique préalable des activités de collecte de données.

## Références — Le contexte décisionnel

Alkin, M.C., *A Guide for Evaluation Decision Makers*, Thousand Oaks : Sage Publications, 1986.

Baird, B.F., *Managerial Decisions under Uncertainty*, New York : Wiley Interscience, 1989.

Cabatoff, Kenneth A., «Getting On and Off the Policy Agenda : A Dualistic Theory of Program Evaluation Utilization», *Canadian Journal of Program Evaluation*,. Vol. 11, N° 2, automne 1996, p. 35 à 60.

Ciarlo, J., éd., *Utilizing Evaluation*, Thousand Oaks : Sage Publications, 1984.

Goldman, Francis et Edith Brashares, «Performance and Accountability : Budget Reform in New Zealand», *Public Budgeting and Finance*, Vol. 11, N° 4, hiver 1991, p. 75 à 85.

Mayne, John et R.S. Mayne, «Will Program Evaluation be Used in Formulating Policy?», in Atkinson, M. et M. Chandler, éd., *The Politics of Canadian Public Policy*, Toronto : University of Toronto Press, 1983.

Moore, M.H., *Creating Public Value : Strategic Management in Government*, Boston : Harvard University Press, 1995.

Nutt, P.C. et R.W. Backoff, *Strategic Management of Public and Third Sector Organizations*, San Francisco : Jossey-Bass, 1992.

O'Brecht, Michael, «Stakeholder Pressures and Organizational Structure», *Canadian Journal of Program Evaluation*, Vol. 7, N° 2, octobre-novembre 1992, p. 139 à 147.

Peters, Guy B. et Donald J. Savoie, Centre canadien de gestion, *Governance in a Changing Environment*, Montréal et Kingston : McGill-Queen's University Press, 1993.

Pressman, J.L. et A. Wildavsky, *Implementation*, Los Angeles : UCLA Press, 1973.

Reavy, Pat, *et al.*, «Evaluation as Management Support : The Role of the Evaluator», *Canadian Journal of Program Evaluation*, Vol. 8, N° 2, octobre-novembre 1993, p. 95 à 104.

Rist, Ray C., éd., *Program Evaluation and the Management of the Government*, New Brunswick (NJ) : Transaction Publishers, 1990.

Schick, Allen, *The Spirit of Reform : Managing the New Zealand State*, rapport commandé par le ministère du Trésor et la Commission des services gouvernementaux de la Nouvelle-Zélande, 1996.

Seidle, Leslie, *Rethinking the Delivery of Public Services to Citizens*, Montréal : Institut de recherches en politiques publiques (IMPP), 1995.

Thomas, Paul G., *The Politics and Management of Performance Measurement and Service Standards*, Winnipeg : St.-John's College, University of Manitoba, 1996.

**Notes :**

### 2.4.3 Stratégies multiples

Une stratégie d'évaluation produit des preuves d'un résultat, tandis qu'une étude d'évaluation porte ordinairement sur plusieurs questions, ce qui signifie qu'on a donc intérêt à faire appel à plusieurs stratégies, d'autant plus qu'il peut aussi être souhaitable d'en utiliser plus d'une pour examiner une question donnée, afin d'accroître l'exactitude et la crédibilité des constatations de l'évaluation.

La plupart des stratégies d'évaluation élaborées pour étudier une question précise peuvent aussi être utilisées pour en examiner d'autres, avec certaines modifications. Même si une stratégie n'est pas idéale pour étudier une autre question, il peut être utile de s'en servir parce que son coût marginal est faible. Supposons par exemple qu'on fasse une étude afin de déterminer l'aptitude à la lecture de deux groupes, dont l'un participant à un programme donné. On fait passer aux membres de chaque groupe un test destiné à mesurer leur aptitude à la lecture, en leur posant aussi diverses questions sur l'utilité et l'efficacité du programme. Les résultats reflètent bien entendu les lacunes inhérentes à tous les résultats des enquêtes sur les attitudes, mais ajoutent quand même des indications aux résultats objectifs du test de lecture, à un coût relativement faible.

La seconde raison d'envisager le recours à plusieurs stratégies de recherche dans une évaluation, c'est qu'il est souvent souhaitable de mesurer ou d'évaluer le même résultat en fonction de plusieurs sources de données, ou en appliquant des modèles d'évaluation différents. En effet, il est souvent difficile, sinon impossible, de mesurer exactement et sans équivoque un résultat donné. Des facteurs de confusion, des erreurs de mesure et des préjugés personnels risquent de se combiner pour saper la validité ou la fiabilité des résultats obtenus lorsqu'on n'a utilisé qu'une seule et unique méthode d'analyse. En effet, les modèles d'évaluation sont habituellement vulnérables à plusieurs obstacles à la validité interne; il est donc impossible d'éliminer ou de tenir compte de toutes les autres explications plausibles. Par conséquent, on doit souvent avoir recours à des stratégies complémentaires pour infirmer les explications indésirables des résultats observés.

C'est pour ces deux raisons qu'il est préférable d'étudier les questions à évaluer de plusieurs points de vue, en se fondant sur plusieurs modalités d'établissement de la preuve afin d'accroître la crédibilité des constatations. Quand des stratégies distinctes qui sont fondées sur des sources de données et des méthodes d'analyse différentes aboutissent à la même conclusion, l'évaluateur peut raisonnablement les considérer comme fiables. Par contre, lorsqu'elles mènent à des conclusions différentes, la situation est évidemment beaucoup moins facile à trancher. Néanmoins, c'est un résultat préférable à ce qui se produit quand on se fonde sur une seule stratégie, en aboutissant sans s'en rendre compte à des conclusions qui pourraient être contradictoires pour peu qu'on en utilise une autre. Lorsque les conclusions diffèrent, c'est peut-être parce que les résultats du programme sont trop sensibles pour pouvoir être mesurés avec précision (ce qui signifie que l'erreur d'échantillonnage l'emporte sur l'effet incrémentiel); pour corriger le problème, il faut alors avoir recours à une meilleure méthode d'analyse ou recueillir plus de données, ou encore à une combinaison de ces deux approches.

Supposons par exemple qu'on tente d'évaluer les effets de notre fameux problème d'aide à un secteur d'activité industrielle. L'évaluation devrait porter sur l'effet incrémentiel du projet, ce qui reviendrait à essayer de déterminer si l'aide fournie a mené à la réalisation du projet envisagé. Cette question pourrait être étudiée sous plusieurs angles différents. Une stratégie consisterait à mener un sondage auprès des cadres des entreprises visées en leur posant la question directement ou indirectement. Cependant, pour diverses raisons, notamment parce qu'ils voudraient obtenir d'autres subventions, les répondants pourraient tendre à exagérer l'effet incrémentiel du programme. Il faudrait donc utiliser d'autres méthodes pour le déterminer. Par exemple, un examen détaillé des registres financiers et de marketing pour la période précédant immédiatement la mise en œuvre du projet permettrait de juger si le rendement attendu des investissements justifiait son exécution sans l'aide de l'État. On pourrait aussi avoir recours à un modèle quasi expérimental avec une analyse correspondante comme nous le verrons au chapitre 3, pour comparer la réalisation de projets non subventionnés à celle de projets qui l'ont été, ou encore pour comparer la fréquence des projets exécutés avant et après la mise en œuvre du programme.

Prenons aussi un autre exemple, celui des enquêtes postales qui peuvent avoir un très vaste rayonnement dans une population cible. Malheureusement, il s'agit là d'une stratégie qui ne se prête généralement pas à des études en profondeur, bien qu'elle puisse être renforcée grâce à des études de cas ou à des entrevues individuelles.

De même, les modèles implicites faisant appel à une analyse du contenu en soi sont peu fiables. Même si ces modèles sont utiles pour l'examen d'avantages difficiles à mesurer, il convient de les compléter par des stratégies plus fiables à fondement quasi expérimental, ce qui augmente énormément la crédibilité globale des constatations de l'évaluation.

### **Références : Stratégies multiples**

Jorjani, Hamid, «The Holistic Perspective in the Evaluation of Public Programs : A Conceptual Framework», *Canadian Journal of Program Evaluation*, Vol. 9, N° 2, octobre-novembre 1994, p. 71 à 92.

### **Notes :**



## 2.5 Résumé

Dans ce chapitre, nous avons analysé les aspects de la recherche et du contexte décisionnel dont il faut tenir compte pour l'élaboration et l'application de méthodes d'évaluation crédibles. Ce faisant, nous avons insisté sur la nécessité de ne jamais négliger les facteurs contextuels inhérents à toutes les études d'évaluation menées dans l'administration fédérale. Ces facteurs sont au moins aussi importants que les questions de recherche qui sont traditionnellement associées à une stratégie d'évaluation.

De plus, le présent chapitre décrit le bien-fondé de multiples éléments probants, soit le recours à plus d'une stratégie d'évaluation pour appuyer les inférences sur les effets du programme. Compte tenu des contraintes temporelles et financières on devrait toujours rechercher de multiples éléments probants pour appuyer les conclusions de l'évaluation.

## Chapitre 3

### MODÈLES D'ÉVALUATION

#### 3.1 Introduction

Un modèle d'évaluation décrit le système logique à appliquer pour recueillir de l'information sur les résultats susceptibles d'être attribués à un programme. La figure 2 illustre le principe fondamental du modèle expérimental, qui implique la comparaison de deux groupes (dont l'un exposé au programme), en attribuant toutes les différences entre les deux groupes au programme lui-même. On appelle ce type de modèle *modèle d'évaluation idéal*. Comme nous l'avons déjà vu, c'est un idéal difficile à atteindre dans la pratique. Pourtant, il est utile aux fins de comparaison et d'explication. On peut l'illustrer de la façon suivante :

	Mesure avant	Exposition au programme	Mesure après
Groupe expérimental	$0_1$	X	$0_3$
Groupe témoin	$0_2$		$0_4$

Dans ce schéma, «0» désigne une mesure ou une observation du résultat du programme et «X», l'exposition au programme. Les chiffres en indices indiquent des mesures ou des traitements différents. Le  $0_1$  représente des estimations (des moyennes estimatives, par exemple) fondées sur les observations relatives à des membres d'un groupe. Il faudrait interpréter des formules comme  $0_3 - 0_4$  comme des indications théoriques, plutôt que comme des écarts entre deux observations. Le schéma montre aussi à quel moment l'observation est faite (avant ou après l'exposition au programme). Nous emploierons la même symbolisation dans tout le chapitre pour illustrer de façon schématique les modèles décrits.

Dans le modèle d'évaluation idéal, le résultat attribué au programme est manifestement  $0_3 - 0_4$ , puisque  $0_1 = 0_2$  et qu'il s'ensuit que  $0_3 = 0_4 + X$  (le programme), ou que  $0_3 - 0_4 = X$ . Remarquons que, dans ce cas-ci, il n'est pas nécessaire que  $0_1$  et  $0_2$  déterminent le résultat net du programme, puisqu'on postule que leurs valeurs sont égales. Il s'ensuit donc que le modèle idéal pourrait être représenté comme suit :

	Exposition au programme	Mesure après
Groupe expérimental	X	$O_3$
Groupe témoin		$O_4$

Il se peut toutefois que l'évaluateur s'intéresse au changement relatif qui s'est produit, auquel cas il doit absolument prendre la mesure avant le programme.

Le modèle idéal est important parce qu'il sert de preuve sous-jacente de l'attribution des résultats à un programme pour tous les modèles d'évaluation décrits dans le présent chapitre. Par exemple, pour faire des inférences causales, il faut *comparer* des groupes identiques, sauf pour l'exposition au programme, avant et après celle-ci. (D'ailleurs, la caractéristique commune de tous les modèles examinés dans ces pages est l'utilisation de la *comparaison*.) Le facteur qui distingue les modèles d'évaluation est le degré auquel on les compare aux groupes en tout point identiques, sauf pour l'exposition au programme.

Dans les modèles les plus rigoureux, appelés **modèles expérimentaux ou aléatoires**, on tente d'assurer l'équivalence initiale des deux groupes en répartissant de façon aléatoire les sujets en deux groupes, un groupe de participants et un groupe témoin. De cette façon, les groupes à comparer s'équivalent, c'est-à-dire que le processus fait en sorte que les valeurs *attendues* (ainsi que les autres caractéristiques de distribution) de  $O_1$  et  $O_2$  soient égales. Nous étudierons les modèles expérimentaux ou aléatoires à la section 3.2.

Les modèles «intermédiaires», dits **quasi expérimentaux**, sont analysés à la section 3.3. Dans ces modèles, qui ressemblent aux modèles expérimentaux en ce sens que des groupes de comparaison servent à faire des inférences causales, on n'a pas recours aux modèles aléatoires ou à la «randomisation» pour créer un groupe de participants (ou expérimental) et un groupe témoin. On part généralement du principe que le groupe de participants est un acquis, ce qui signifie qu'on choisit un ou des groupes de comparaison (ou témoins) de façon qu'ils y correspondent le plus étroitement possible. Lorsqu'il n'y a pas randomisation, il n'est plus possible de postuler la comparabilité des groupes, de sorte qu'il faut trouver des moyens de remédier à leur éventuelle incomparabilité. Néanmoins, les modèles quasi expérimentaux demeurent les meilleurs lorsque la randomisation n'est pas possible.

On trouve à l'autre extrémité de l'échelle les **modèles implicites**, qui se prêtent habituellement mal à la mesure des changements et à leur attribution à un programme. En voici un exemple :

	Exposition au programme	Mesure après
Groupe expérimental	X	O <sub>1</sub>

Dans un modèle comme celui-là, on prend une mesure après l'exposition au programme, en posant des hypothèses sur les conditions présentes avant sa mise en œuvre. On postule que tous les changements par rapport à la situation existante avant le programme lui sont attribuables. Autrement dit, on part de l'hypothèse qu'on ne constaterait aucun changement à l'égard d'un groupe témoin non précisé (ou du moins pas de changements de l'ampleur de ceux qui sont constatés pour le groupe expérimental). Nous reviendrons plus longuement sur les modèles implicites à la section 3.4.

Ces types de modèles ont tous des degrés différents de rigueur quant à l'établissement des résultats d'un programme; ils traduisent aussi une différence fondamentale entre les programmes expérimentaux et les programmes «ordinaires», c'est-à-dire non expérimentaux. Or, la plupart des programmes gouvernementaux ont pour objet d'apporter des avantages aux participants, et les administrateurs partent du principe qu'ils sont bel et bien efficaces. La participation à ces programmes est typiquement déterminée en fonction de critères d'admissibilité. C'est bien différent dans le cas des programmes expérimentaux ou pilotes, qui sont mis en œuvre pour vérifier la validité théorique d'un programme et pour en déterminer l'efficacité. Les participants aux programmes de ce genre en retirent des avantages, mais c'est un résultat secondaire, le but étant essentiellement de vérifier si les programmes sont efficaces. Il s'ensuit que les participants sont souvent choisis pour maximiser les chances d'obtention de résultats concluants, pas nécessairement en fonction de critères d'admissibilité.

Ces deux buts, à savoir générer des avantages et vérifier la validité de la théorie sur laquelle le programme est fondé, sont presque toujours incompatibles. Les gestionnaires des programmes estiment normalement que ceux-ci ont pour objet de produire des avantages, même dans le cas d'un programme pilote. Les évaluateurs et les planificateurs, d'autre part, préfèrent exécuter un programme expérimental pour déterminer d'emblée s'il vaut la peine de lui donner plus d'importance. Dans la pratique, la plupart des programmes ne sont pas mis à l'essai, ce qui signifie que l'évaluateur doit fréquemment opter pour des modèles d'évaluation non expérimentaux.

Dans le présent chapitre, nous allons analyser les trois types de modèles d'évaluation dont nous venons de faire état. Nous allons décrire des modèles de chaque type en précisant leurs avantages et leurs inconvénients. Nous répartissons les modèles en trois types — aléatoires, quasi expérimentaux et implicites — pour faciliter l'analyse, mais la distinction entre les trois n'est pas toujours rigoureuse. En effet, les modèles quasi expérimentaux se confondent souvent avec les modèles implicites. Néanmoins, les distinctions sont utiles et révèlent, dans la plupart des cas,

un degré de rigueur différent. L'évaluateur qui passe du modèle aléatoire au modèle implicite doit tenir compte d'un nombre croissant d'obstacles à la validité de ses inférences causales.

### **Références : Modèles d'évaluation**

Abt, C.G., éd., *The Evaluation of Social Programs*, Thousand Oaks : Sage Publications, 1976.

Boruch, R.F., «Conducting Social Experiments», *Evaluation Practice in Review*, Vol. 34 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1987.

Campbell, D.T. et J.C. Stanley, *Experimental and Quasi-experimental Designs for Research*, Chicago : Rand-McNally, 1963.

Cook, T.D. et D.T. Campbell, *Quasi-experimentation: Designs and Analysis Issues for Field Settings*, Chicago : Rand-McNally, 1979.

Datta, L. et R. Perloff, *Improving Evaluations*, Thousand Oaks : Sage Publications, 1979, section II.

Globerson, Aryé, et al., *You Can't Manage What You Don't Measure: Control and Evaluation in Organizations*, Brookfield : Gower Publications, 1991.

Rossi, P.H. et H.E. Freeman, *Evaluation: A Systematic Approach* (2<sup>e</sup> édition), Thousand Oaks : Sage Publications, 1989.

Trochim, W.M.K., éd., *Advances in Quasi-experimental Design and Analysis*, Vol. 31 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1986.

Watson, Kenneth, «Program Design Can Make Outcome Evaluation Impossible: A Review of Four Studies of Community Economic Development Programs», *Canadian Journal of Program Evaluation*, Vol. 10, N<sup>o</sup> 1, avril-mai 1995, p. 59 à 72.

Weiss, C.H., *Evaluation Research*, Englewood Cliffs (NJ) : Prentice-Hall, 1972, chapitre 4.

### **Notes :**

## 3.2 Modèles expérimentaux aléatoires

L'approche la plus rigoureuse pour établir des relations causales entre un programme et ses résultats est celle des modèles expérimentaux. S'ils sont bien appliqués, ces modèles fournissent les preuves les plus concluantes des effets du programme. Malheureusement, pour bien des programmes gouvernementaux, ils sont impossibles à mettre en œuvre, pour peu que le programme soit offert depuis un certain temps. Néanmoins, ils sont importants, et ce pour les deux raisons suivantes.

Premièrement, ils sont aussi près que possible du modèle d'évaluation idéal que nous venons de décrire. Par conséquent, même s'il n'est pas possible d'appliquer un modèle expérimental, les modèles moins rigoureux sont souvent cotés selon leur degré de conformité à ce modèle expérimental, et c'est pour cette raison qu'il est important de comprendre leurs avantages et leurs inconvénients.

Deuxièmement, malgré leurs difficultés d'ordre pratique, les modèles expérimentaux peuvent être utilisés pour évaluer de nombreux programmes; ils l'ont souvent été, d'ailleurs. Par exemple, on s'est servi d'un modèle expérimental pour évaluer les programmes scolaires conçus afin de prévenir la consommation et l'abus d'alcool chez les adolescents, en faisant appel à un groupe expérimental et à un groupe témoin (certaines classes ayant accès au programme et d'autres pas) pour obtenir des mesures de l'attitude, des connaissances, des intentions et de la consommation réelle de boissons alcooliques (Schlegel, 1977).

Les modèles expérimentaux ou aléatoires sont caractérisés par la répartition aléatoire des participants éventuels entre le groupe expérimental et le groupe témoin, afin d'assurer l'équivalence des deux. On dit qu'il s'agit d'expériences, en ce sens que les participants au programme sont choisis au hasard parmi tous les candidats possibles. Il existe un grand nombre de modèles expérimentaux, dont les quatre suivants :

- modèle aléatoire classique avec groupe témoin;
- modèle aléatoire avec mesure après le programme seulement et groupe témoin;
- modèle avec blocs aléatoires et carré latin;
- modèle factoriel.

Il ne faut pas oublier que l'expression *modèle aléatoire* n'est pas synonyme d'*échantillonnage aléatoire*. Dans le premier cas, on choisit au hasard des membres d'une population cible soit pour le groupe témoin, soit pour le groupe expérimental, tandis que dans le second, on se fonde sur un calcul des probabilités pour choisir un échantillon d'une population donnée. L'échantillonnage aléatoire fondé sur deux populations différentes ne générerait pas de groupes équivalents aux fins d'une évaluation expérimentale.

### Modèle aléatoire classique avec groupe témoin

On entend par là un modèle expérimental classique, qui peut être représenté de la façon suivante, «R» désignant une répartition aléatoire :

	Mesure avant	Exposition au programme	Mesure après
Groupe expérimental (R)	$0_1$	X	$0_3$
Groupe témoin (R)	$0_2$		$0_4$

Dans ce modèle, les participants éventuels au programme qui font partie de la population cible sont choisis au hasard, pour être répartis dans le groupe expérimental (où ils sont exposés au programme) ou le groupe témoin. On prend des mesures avant et après le programme (mesures préalables et postérieures au programme), dont le résultat net est représenté schématiquement par la formule  $(0_3 - 0_4) - (0_1 - 0_2)$ .

La répartition aléatoire (ou, si l'on préfère, la randomisation) signifie qu'il y a pour chaque membre de la population cible une probabilité connue qu'il soit choisi pour faire partie du groupe expérimental ou du groupe témoin. Ces probabilités sont souvent égales, auquel cas chaque membre a des chances égales d'être choisi pour faire partie d'un groupe ou de l'autre. Par suite de cette randomisation, les groupes expérimental et témoin sont mathématiquement équivalents, ce qui signifie que les valeurs attendues de  $0_1$  et  $0_2$  sont égales. Toutefois, les mesures réelles prises avant le programme peuvent varier de façon aléatoire, et c'est pourquoi elles peuvent donner une meilleure idée du résultat net, puisqu'elles permettent de tenir compte de toute différence susceptible d'exister entre les groupes ( $0_1 - 0_2$ ) en dépit de la randomisation. Bref, dans ce modèle, l'intervention du programme (ou le traitement) est la seule différence, à part le hasard, entre le groupe expérimental et le groupe témoin.

### Modèle aléatoire avec mesure après le programme seulement et groupe témoin

Le modèle aléatoire classique a notamment pour inconvénient d'être vulnérable à une distorsion attribuable à l'essai. La validité de l'évaluation est en effet menacée puisque la mesure préalable au programme elle-même peut influencer sur le comportement du groupe expérimental ou du groupe témoin (ou des deux), et ce à tel point que toute inférence causale que l'évaluateur pourrait vouloir faire risquerait d'être mise en doute. Pour éviter cette difficulté, l'évaluateur peut décider de ne pas faire de mesure avant le programme, auquel cas le schéma de son modèle se présente comme suit :

	Exposition au programme	Mesure après
Groupe expérimental (R)	X	$O_1$
Groupe témoin (R)		$O_2$

Le modèle aléatoire avec mesure après le programme peut être extrêmement rigoureux. Toutefois, il faut bien se rappeler que, même avec une répartition aléatoire, il se peut que les deux groupes choisis soient nettement différents quant aux mesures d'intérêt. On n'est donc jamais sûr d'avoir complètement éliminé les différences entre les groupes initiaux susceptibles d'influer sur le résultat de l'évaluation.

### Modèle avec blocs aléatoires et carré latin

Afin de minimiser la probabilité que l'effet net d'un programme soit imputable à une erreur d'échantillonnage, il est préférable d'utiliser un échantillon aussi gros que possible. Malheureusement, cela peut coûter fort cher. Pour éviter cet écueil, on pourrait combiner la répartition aléatoire et l'*appariement des sujets (constitution de blocs)* lorsqu'il faut absolument utiliser des échantillons relativement petits. L'appariement consiste à diviser la population cible dans laquelle les membres du groupe expérimental et du groupe témoin sont choisis en «blocs» définis en fonction d'une ou de plusieurs variables qui devraient influencer sur les résultats du programme.

Par exemple, si l'on s'attend à ce que les urbains réagissent plus favorablement que les ruraux à un programme social, on peut constituer deux blocs, l'un urbain et l'autre rural. Ensuite, à l'intérieur de chaque bloc, on fait une répartition aléatoire pour choisir les membres du groupe expérimental et du groupe témoin. Cette approche pourrait contribuer à assurer une participation raisonnablement égale des sujets urbains et ruraux. En fait, on devrait toujours opter pour l'appariement des sujets lorsque les variables importantes sont connues.

Bien entendu, les groupes peuvent être appariés en fonction de plus d'une variable. Néanmoins, l'augmentation du nombre de variables fait vite augmenter le nombre de blocs et, par conséquent, la taille de l'échantillon nécessaire. Ainsi, quand on s'attend à ce que la langue officielle parlée (le français ou l'anglais) influe sur les résultats du programme, il faut envisager de créer les blocs suivants : urbain anglophone, rural anglophone, urbain francophone et rural francophone. En outre, puisque chaque bloc doit contenir un groupe expérimental et un groupe témoin, il faut en constituer huit en tout, en respectant pour chacun les règles relatives à la taille minimale de l'échantillon. Heureusement, il existe des méthodes pour réduire le nombre de groupes nécessaires, comme le modèle du carré latin. Ces méthodes ne peuvent toutefois être employées que si les effets d'interaction entre les variables du groupe expérimental et du groupe témoin sont relativement minimes.



## Modèle factoriel

Dans les modèles classiques et dans ceux qui font appel à des blocs aléatoires, il n'y a qu'une seule variable expérimentale (ou de traitement) en jeu. Or, les programmes font souvent appel à toute une série d'incitations pour aiguiller les bénéficiaires vers un résultat recherché. Quand l'évaluateur est en mesure de distinguer les effets des différentes méthodes d'intervention utilisées, il peut avoir recours à un modèle factoriel, ce qui lui permet non seulement de distinguer les effets particuliers de chaque variable expérimental, mais aussi d'estimer les effets nets mixtes (les effets d'interaction) de paires de variables expérimentales. C'est un grand avantage, puisqu'on observe souvent des effets d'interaction dans les phénomènes sociaux. Par exemple, les effets combinés d'une hausse des taxes sur le tabac et d'une augmentation du budget anti-tabagisme peuvent être plus marqués que la somme des effets isolés des deux.

## Avantages et inconvénients

Les modèles expérimentaux sont les plus rigoureux lorsqu'il s'agit de faire des inférences causales sur les résultats des programmes. Ils permettent en effet d'éliminer la plupart des obstacles à la validité interne, puisqu'ils font appel à un groupe témoin, à des modèles aléatoires ainsi qu'à des modèles avec blocs aléatoires et à des modèles factoriels. Leur principal inconvénient est qu'ils sont souvent difficiles à appliquer.

Malheureusement, la randomisation (c.-à-d. la répartition aléatoire entre le groupe expérimental et le groupe témoin) n'est souvent pas possible :

- quand toute la population cible bénéficie du programme, il n'existe aucune variable pouvant servir à la constitution d'un groupe témoin;
- quand le programme fonctionne depuis assez longtemps, il risque probablement d'exister des différences sensibles entre ceux qui y ont participé (le groupe expérimental éventuel) et les autres (le groupe témoin éventuel);
- il peut être illégal ou immoral de faire bénéficier du programme certaines personnes (les membres du groupe expérimental) en en privant d'autres (les membres du groupe témoin).

La plupart des programmes gouvernementaux correspondent manifestement à au moins un des cas qui précèdent, ce qui fait que la randomisation est extrêmement difficile à leur égard, sauf peut-être s'ils sont considérés comme une véritable expérience, comme dans un programme pilote.

Les modèles expérimentaux sont toujours vulnérables à tous les obstacles à la validité externe et à certains de ceux qui sapent la validité interne.

Même avec un modèle expérimental, la difficulté de généraliser à partir des conclusions sur les résultats d'un programme n'est pas automatiquement éliminée. Par exemple, la randomisation aux fins de généralisation est une tout autre question que la sélection aléatoire des groupes expérimental et témoin, car elle exige que la population cible initiale à partir de laquelle les deux groupes seront créés soit elle-même choisie au hasard à même la population des bénéficiaires éventuels en général (soit la population des sujets à l'égard desquels l'évaluateur peut souhaiter généraliser ses résultats).

En outre, plusieurs obstacles importants à la validité interne perdurent, malgré le choix au hasard des membres des deux groupes :

- l'*attrition différentielle* (ou le retrait des membres des groupes expérimental et témoin) pourrait fausser la randomisation initiale;
- la *diffusion du traitement* entre les deux groupes pourrait fausser les résultats.

De plus, le modèle expérimental classique présente lui aussi des risques :

- des *changements d'instruments* pourraient de toute évidence fausser les mesures prises;
- la *réaction à l'essai* pourrait entraîner des comportements différents des membres du groupe expérimental et du groupe témoin.

Comme ces deux derniers facteurs sont essentiellement attribuables aux essais préalables, le modèle aléatoire avec mesure après le programme seulement en comparaison avec le groupe témoin peut les éviter, comme nous l'avons déjà expliqué. Il faudrait néanmoins comprendre clairement que, en dépit de leurs avantages, les résultats obtenus grâce aux modèles expérimentaux devraient être interprétés avec beaucoup de circonspection.

### Références : Modèles expérimentaux aléatoires

Boruch, R.F., «Conducting Social Experiments», *Evaluation Practice in Review*, Vol. 34 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1987, p. 45 à 66.

Boruch, R.F., «On Common Contentions About Randomized Field Experiments», in Gene V. Glass, éd., *Evaluation Studies Review Annual*, Thousand Oaks : Sage Publications, 1976.

Campbell, D., «Considering the Case Against Experimental Evaluations of Social Innovations», *Administrative Science Quarterly*, Vol. 15, N° 1, 1970, p. 111 à 122.

Eaton, Frank, «Measuring Program Effects in the Presence of Selection Bias: The Evolution of Practice», *Canadian Journal of Program Evaluation*, Vol. 9, N° 2, octobre-novembre 1994, p. 57 à 70.

Trochim, W.M.K., éd., «Advances in Quasi-experimental Design and Analysis», Vol. 31 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1986.

**Notes :**

### 3.3 Modèles quasi expérimentaux

Même lorsqu'on est incapable de faire une randomisation, il peut être possible d'établir un groupe témoin ressemblant suffisamment au groupe expérimental pour permettre des inférences valides sur les résultats attribuables au programme. Dans la présente section, on entend par «modèles quasi expérimentaux» ceux pour lesquels on a recours à un groupe témoin non aléatoire pour faire des inférences sur les résultats d'un programme. Le groupe témoin pourrait être soit un *groupe créé* de toutes pièces qui n'a pas été exposé au programme, soit un *groupe réflexif*, c'est-à-dire le groupe expérimental lui-même avant son exposition au programme.

Nous allons décrire trois modèles quasi expérimentaux généraux, à savoir :

- les modèles à mesures avant et après le programme;
- les modèles à série temporelle ou modèles chronologiques;
- les modèles à mesures prises après le programme seulement.

Les trois modèles sont présentés en ordre de rigueur descendant, même si le degré d'équivalence entre le groupe expérimental et le groupe témoin est le facteur déterminant de la rigueur du modèle dans chaque cas.

#### 3.3.1 Modèles dont les mesures sont prises avant et après le programme

Il y a fondamentalement deux types de modèles de ce genre, ceux dont les mesures sont prises avant et après le programme avec groupe témoin non équivalent et ceux dont les mesures sont prises avant et après le programme avec un seul groupe, le groupe expérimental. Dans le premier cas, on utilise un groupe témoin créé; dans le second, un groupe témoin réflexif.

##### Modèles dont les mesures sont prises avant et après le programme avec groupe témoin non équivalent

Ce modèle, dont la structure est analogue à celle du modèle expérimental classique, est fondé sur des mesures prises avant et après le programme dans le groupe expérimental et dans un groupe témoin.

	Mesure avant	Exposition au programme	Mesure après
Groupe expérimental	$0_1$	X	$0_3$
Groupe témoin	$0_2$		$0_4$

Le groupe témoin est choisi de façon que ses caractéristiques importantes ressemblent le plus possible à celles du groupe expérimental. Le degré de similarité entre les groupes est déterminé grâce à une comparaison effectuée avant le programme. Dans la mesure où l'on a réalisé un bon appariement (fondé sur les variables qui semblent exercer une influence sur celles des résultats), on peut dire que ce modèle s'apparente au modèle aléatoire avec groupe témoin, et qu'il permet de minimiser les obstacles à la validité interne. Malheureusement, il est généralement difficile d'apparier parfaitement toutes les variables importantes, de sorte qu'il subsiste normalement au moins une autre explication plausible des résultats observés nets du programme, à savoir que les deux groupes n'étaient pas égaux au départ.

### Modèle dont les mesures sont prises avant et après le programme avec un seul groupe

On utilise souvent ce modèle simple en dépit de toute ses lacunes intrinsèques, probablement parce que c'est celui qui ressemble le plus à ce qu'on entend communément par les résultats d'un programme, à savoir les changements survenus entre la période antérieure au programme et celle qui lui est postérieure. On peut le représenter de la façon suivante :

	Mesure avant	Exposition au programme	Mesure après
Groupe expérimental	$O_1$	X	$O_2$

De nombreux obstacles peuvent saper la validité interne de ce modèle, car bien des explications plausibles pourraient justifier les différences constatées entre  $O_2$  et  $O_1$ , puisque le groupe témoin est en l'occurrence le groupe expérimental avant son exposition au programme (c'est un groupe témoin réflexif). L'absence de groupe témoin distinct signifie que la plupart des obstacles à la validité interne sont présents. Les *événements historiques* risquent d'ailleurs de poser un problème, étant donné que le modèle ne peut tenir compte des événements extérieurs au programme qui influent sur les résultats observés. La *maturation* normale de la population visée peut elle-même expliquer les changements, le cas échéant. En outre, le changement observé peut être simplement un *facteur de régression*, et  $O_1$  peut être anormalement faible, de sorte que la mesure  $O_2 - O_1$  porte davantage sur une fluctuation aléatoire que sur un changement attribuable au programme. Enfin, les *essais*, les *instruments* et l'*attrition* peuvent tous poser des problèmes.

Le seul avantage de ce modèle est sa simplicité. Si l'évaluateur réussit à tenir suffisamment compte des facteurs externes, le modèle fournit de l'information et des données raisonnablement valides et concluantes. Dans le domaine des sciences naturelles, on arrive habituellement à contrôler suffisamment les facteurs externes, en laboratoire, mais c'est beaucoup plus difficile dans le domaine des sciences sociales.

### 3.3.2 Modèles à série temporelle ou modèles chronologiques

Les modèles à série temporelle ou modèles chronologiques sont caractérisés par une série de mesures échelonnées dans le temps à la fois avant et après l'exposition au programme. Tous les modèles que nous avons décrits dont les mesures sont prises avant et après le programme pourraient être transformés en modèles chronologiques. Autrement dit, les modèles chronologiques pour lesquels il n'existe que quelques mesures avant et après le programme sont vulnérables aux mêmes obstacles influant sur la validité interne que les modèles à mesures uniques correspondants. À l'inverse, une série complète de mesures avant et après le programme permet à l'évaluateur d'éliminer un grand nombre de ces obstacles, en analysant les tendances antérieures et postérieures au programme.

Nous allons maintenant décrire deux modèles chronologiques :

- le modèle de base à série temporelle
- le modèle à série temporelle avec groupe témoin non équivalent.

#### Modèle de base à série temporelle

Le modèle de base à série temporelle est un modèle chronologique courant grâce auquel on peut prendre un nombre quelconque de mesures antérieures et postérieures au programme. Il peut être représenté comme suit :

	Mesure avant	Exposition au programme	Mesure après
Groupe expérimental	0 <sub>1</sub> 0 <sub>2</sub> 0 <sub>3</sub> 0 <sub>4</sub>	X	0 <sub>5</sub> 0 <sub>6</sub> 0 <sub>7</sub> 0 <sub>8</sub>

Avec ce modèle, l'évaluateur peut déterminer les effets d'un programme donné en fonction du changement qui se manifeste dans la série de mesures prises avant et après l'exposition au programme. Si les données de la série temporelle sont fiables, le modèle peut être relativement rigoureux, auquel cas il permet d'éliminer de nombreux obstacles à la validité interne, notamment les effets de *maturation* et d'*essai*. Certains autres obstacles subsistent quand même, notamment ceux relatifs aux *événements historiques*, parce que les modèles à série temporelle ne peuvent éliminer le risque qu'un facteur autre que le programme ait produit le changement entre le moment où les mesures ont été prises avant le programme et celui où elles l'ont été après.

## Modèles à série temporelle avec groupe témoin non équivalent

Les modèles à série temporelle peuvent être améliorés lorsqu'on y ajoute des groupes témoins, comme dans le modèle à série temporelle avec groupe témoin non équivalent ci-dessous :

	Mesure avant	Exposition au programme	Mesure après
Groupe expérimental	0 <sub>1</sub> 0 <sub>2</sub> 0 <sub>3</sub> 0 <sub>4</sub> 0 <sub>5</sub>	X	0 <sub>11</sub> 0 <sub>12</sub> 0 <sub>13</sub> 0 <sub>14</sub> 0 <sub>15</sub>
Groupe témoin	0 <sub>6</sub> 0 <sub>7</sub> 0 <sub>8</sub> 0 <sub>9</sub> 0 <sub>10</sub>		0 <sub>16</sub> 0 <sub>17</sub> 0 <sub>18</sub> 0 <sub>19</sub> 0 <sub>20</sub>

Puisque le groupe expérimental et le groupe témoin devraient normalement être soumis aux mêmes facteurs externes, il est peu probable qu'un changement observé soit attribuable à un autre facteur que le programme. Comme pour n'importe quel modèle dans lequel on utilise un groupe témoin non équivalent, il faut toutefois que les groupes se ressemblent suffisamment en ce qui concerne les caractéristiques étudiées. Si c'est le cas, un modèle chronologique comme celui-ci peut se révéler très rigoureux.

Il faut quand même signaler un certain nombre des avantages et des inconvénients de ces modèles.

Les modèles chronologiques fondés sur des données de série temporelle fiables peuvent éliminer de nombreux obstacles à la validité interne.

Cette caractéristique est attribuable au fait que, lorsqu'ils sont bien exécutés, ces modèles rendent possibles une certaine évaluation de la tendance de maturation avant l'intervention du programme.

Les modèles chronologiques peuvent être utilisés pour analyser divers effets du programme dépendant du facteur temps.

L'aspect longitudinal des modèles historiques permet à l'évaluateur qui s'en sert d'analyser plusieurs questions en déterminant, par exemple, si l'effet observé est continu ou s'il s'estompe avec le temps et s'il est immédiat ou à retardement, ou encore saisonnier. Chaque fois qu'une question de ce genre est importante, il faut utiliser un modèle comme celui-là.

On n'a pas toujours de données fiables pour mener l'analyse à série temporelle qui s'impose.

Les modèles chronologiques posent de nombreux problèmes de données. Par exemple, les séries temporelles utilisables sont souvent plus courtes que celles qu'on recommande normalement pour l'analyse statistique (il n'y a pas suffisamment de données); on peut en outre avoir utilisé différentes méthodes de collecte des données au cours de la période à l'étude, et il se peut aussi que les indicateurs aient changé avec le temps.

Lorsqu'on utilise un modèle chronologique, il faut ordinairement faire une analyse spéciale des séries temporelles.

Les régressions des moindres carrés les plus courantes ne se prêtent pas à l'analyse des séries temporelles. Il faut donc recourir à diverses techniques spécialisées (voir par exemple Cook et Campbell, 1979, chapitre 6; Fuller, 1976; Jenkins, 1979; et Ostrom, 1978).

### 3.3.3 Modèles dont les mesures sont prises après le programme seulement

Dans le cas de ces modèles, les mesures sont prises uniquement après l'exposition au programme, ce qui élimine les obstacles associés aux *essais* et aux *instruments*. Néanmoins, puisqu'il n'existe pas d'information sur la situation antérieure au programme, d'importants obstacles à la validité subsistent, même lorsqu'on utilise un groupe témoin. Nous allons décrire deux modèles de ce genre.

#### Modèle dont les mesures sont prises après le programme seulement avec groupe témoin non équivalent

Ce genre de modèle se présente comme suit :

	Exposition au programme	Mesure après
Groupe expérimental	X	O <sub>1</sub>
Groupe témoin		O <sub>2</sub>

La sélection et l'attrition sont les principaux obstacles à la validité interne d'un modèle de ce genre. Il est absolument impossible de savoir si les deux groupes étaient équivalents avant l'exposition au programme. Il se pourrait donc que l'écart entre O<sub>1</sub> et O<sub>2</sub> soit simplement le reflet de leur différence initiale et ne soit donc pas attribuable à l'exposition au programme. Qui plus est, on ne connaît pas le taux d'abandon du programme (*effet attribuable à l'attrition*), faute d'avoir pris des mesures avant son exécution. Enfin, même si les deux groupes avaient été équivalents au départ, il est



possible que  $O_1$  ou  $O_2$  n'incluent pas les personnes qui ont abandonné le programme, ce qui risque d'entraîner une distorsion des estimations de ses effets.

### **Modèle dont les mesures sont prises après le programme seulement avec traitements différents**

C'est un modèle plus juste, qui se présente de la façon suivante :

	Exposition au programme	Mesure après
Groupe 1	$X_1$	$O_1$
Groupe 2	$X_2$	$O_2$
Groupe 3	$X_3$	$O_3$
Groupe 4	$X_4$	$O_4$

Dans ce cas-ci, différents groupes bénéficient du programme à des degrés différents, ce qui peut arriver lorsqu'il y a des variantes régionales de la prestation et des avantages d'un programme national. Si les échantillons sont suffisamment importants, on pourrait faire une analyse statistique pour établir le lien entre les différents niveaux d'application du programme et les résultats observés ( $O_1$ ), tout en tenant compte aussi des autres variables.

Pour ce modèle comme pour le précédent, la sélection et l'attrition sont les principaux obstacles à la validité interne.

### **Avantages et inconvénients**

Il faut de la créativité et du talent pour concevoir un modèle quasi expérimental, qui peut toutefois générer des constatations très précises.

Souvent, il n'y a rien de mieux pour faire une évaluation que d'utiliser un modèle quasi expérimental. En effet, si la randomisation ne permet pas d'établir l'équivalence du groupe expérimental et du groupe témoin, la meilleure solution consiste à exploiter toutes ses connaissances préalables pour choisir le modèle expérimental le moins entaché de facteurs de confusion. D'ailleurs, un modèle quasi expérimental bien exécuté peut aboutir à des constatations plus fiables que celles d'un modèle expérimental mal appliqué.

Les modèles quasi expérimentaux peuvent être moins coûteux et plus faciles à appliquer que les modèles expérimentaux.

Puisque les modèles quasi expérimentaux n'exigent pas le traitement au hasard et un groupe témoin, leur utilisation peut être moins coûteuse et leur application plus facile que celles des modèles expérimentaux.

Lorsqu'on utilise un modèle quasi expérimental, il faut tenir compte individuellement de chaque obstacle à la validité interne.

La mesure dans laquelle les obstacles à la validité interne posent un problème est largement fonction de celle dans laquelle l'évaluateur réussit à appairer le groupe expérimental et le groupe témoin. S'il réussit à définir et à appairer convenablement les principales variables à étudier, il peut réduire énormément les obstacles à la validité interne. Malheureusement, il est souvent impossible d'appairer toutes ces variables-là.

Pour choisir le modèle approprié, l'évaluateur devrait examiner les modèles quasi expérimentaux envisageables, déterminer les principaux obstacles à la validité de chacun et choisir celui qui lui permet d'éliminer ou de réduire le plus les principaux obstacles, ou au moins de tenir compte de leur impact.

**Notes :**

### 3.4 Modèles implicites

Les modèles implicites sont probablement ceux qu'on utilise le plus souvent, mais ce sont aussi les moins rigoureux. Souvent, il est impossible d'en tirer des conclusions fiables. Par contre, ils peuvent s'imposer dans les cas où l'on peut soutenir logiquement qu'un résultat est attribuable au programme. Au fond, ce sont des modèles dont les mesures sont prises après le programme, sans groupe témoin. Schématiquement, on peut les représenter comme suit :

	Exposition au programme	Mesure après
Groupe expérimental	X	$O_1$

Comme on peut le constater dans cet exemple, l'envergure des effets du programme est inconnue (puisque'il n'y a aucune mesure prise avant), et il est impossible d'arriver à des constatations manifestes au sujet de l'attribution ( $O_1$  pourrait être attribuable à une foule de facteurs). Dans la pire des éventualités, cela suppose qu'on demande aux participants s'ils ont «aimé» le programme. Les témoignages positifs sont alors présentés comme preuves de son succès. Campbell (1977), entre autres, déplore cette méthode d'évaluation pourtant très répandue.

Bien que sa popularité soit en partie attribuable à une mauvaise conception de l'évaluation, il arrive parfois que ce modèle soit le seul utilisable, lorsqu'il n'existe pas de mesures antérieures au programme, ni de groupe témoin. En pareil cas, il vaudrait mieux tirer le meilleur parti de la situation en convertissant le modèle implicite en un modèle implicite quasi expérimental, auquel cas il y a trois possibilités :

- modèle avec groupe témoin théorique,
- modèle avec mesures antérieures au programme rétrospectives,
- modèle d'estimation directe de la différence.

Nous allons maintenant décrire ces trois modèles.

#### **Modèle dont les mesures sont prises après le programme seulement avec groupe témoin théorique**

Ce modèle, pour lequel on postule l'équivalence d'un groupe témoin théorique, ressemble à un modèle quasi expérimental dont les mesures sont prises après le programme seulement avec groupe témoin non équivalent. Il se présente de la façon suivante :

	Exposition au programme	Mesure après
Groupe expérimental	X	$O_1$
Groupe témoin théorique		$O_2^*$

La différence, c'est que la mesure  $O_2^*$  est postulée plutôt qu'observée. L'évaluateur pourrait théoriquement postuler que le résultat serait inférieur à un certain niveau si le programme n'avait pas existé. Par exemple, dans le cas d'un programme conçu pour sensibiliser la population aux effets nocifs de la caféine, on pourrait supposer que les connaissances de la Canadienne ou du Canadien moyen ( $O_2^*$ ) seraient négligeables en l'absence d'un programme d'information nationale. Prenons un autre exemple : la détermination de l'avantage économique d'un programme ou d'un projet de l'État. En l'absence d'un programme, on postule souvent que l'investissement équivalent laissé à l'initiative du secteur privé aurait un taux de rendement social moyen de 10 p. 100, soit le  $O_2^*$  dans ce cas, ce qui signifie qu'on comparerait alors le taux de rendement du projet d'investissement gouvernemental ( $O_1$ ) à la norme de 10 p. 100 du secteur privé ( $O_2^*$ ).

### **Modèle dont les mesures sont prises après le programme seulement avec mesures antérieures rétrospectives**

Dans ce cas, on obtient *bel et bien* des mesures avant le programme, quoiqu'*après* l'exposition, de sorte que le modèle ressemble à un modèle quasi expérimental dont les mesures sont prises avant et après le programme :

	Mesure avant (rétrospective)	Exposition au programme	Mesure après
Groupe expérimental	$O_1$	X	$O_2$

Par exemple, supposons que les deux questions suivantes soient posées à des étudiants ayant suivi un cours de français :

1. Sur une échelle de 1 à 5, évaluez votre connaissance du français avant le cours.
2. Sur une échelle de 1 à 5, évaluez votre connaissance du français après le cours.

On demanderait donc aux étudiants d'évaluer leurs connaissances du français avant et après le cours, une fois celui-ci terminé. La différence entre les deux évaluations pourrait servir à déterminer l'efficacité du programme.

### **Modèle dont les mesures sont prises après le programme seulement avec estimation de la différence**

Ce modèle implicite est le moins efficace de tous. On peut le représenter de la façon suivante :

	Exposition au programme	Mesure après
Groupe expérimental	X	$O = (O_2 - O_1)$

Dans ce cas-ci, le répondant estime directement l'effet incrémentiel du programme. On pourrait, par exemple, demander à des représentants d'entreprises combien d'emplois ont été créés grâce à une subvention, ou encore inviter des étudiants qui ont suivi un cours de français à donner une estimation de la nature et de l'étendue des connaissances qu'ils ont acquises grâce au cours. La différence entre ce modèle et celui qui fait appel à des mesures antérieures rétrospectives, c'est que les répondants eux-mêmes doivent répondre directement à la question sur l'effet du programme.

### **Avantages et inconvénients**

Les modèles implicites sont souples, polyvalents et faciles à appliquer.

Puisqu'ils sont peu exigeants, les modèles implicites sont toujours réalisables. En effet, on peut toujours demander aux participants à un programme, aux gestionnaires ou à des spécialistes leur opinion sur ses résultats. Toutefois, cette facilité d'application même peut constituer un inconvénient en ce sens qu'on risque d'être tenté d'opter pour des modèles implicites «faciles», alors qu'on aurait pu avoir recours à un modèle implicite plus rigoureux, voire à un modèle quasi expérimental, avec un peu plus de travail et d'ingéniosité.

Les modèles implicites peuvent servir à étudier virtuellement n'importe quelle question et peuvent être utilisés comme instruments d'exploration.

On peut poser n'importe quelle question sur le programme aux participants ou aux gestionnaires. Malgré leurs lacunes évidentes en ce qui concerne l'examen objectif des résultats du programme et leur attribution, les modèles implicites peuvent entièrement permettre de trouver la réponse à des questions sur l'exécution d'un programme. Dans le cas d'un programme de services, par exemple, ils permettent à

l'évaluateur d'étudier les questions relatives au degré de satisfaction de la clientèle. En outre, avec une enquête ultérieure au programme, ils peuvent cerner un certain nombre de résultats susceptibles d'être étudiés grâce à d'autres stratégies d'évaluation.

Les modèles implicites produisent peu de preuves objectives des résultats découlant d'un programme.

Il est possible de tirer des conclusions sur les résultats d'un programme à l'aide d'un modèle implicite seulement si l'on pose des hypothèses majeures sur ce qui se serait produit en l'absence du programme. Les obstacles à la validité interne sont donc aussi nombreux qu'importants (au titre des *événements historiques*, de la *maturation* et de l'*attrition*, par exemple), et il faut les éliminer un à un.

Lorsque l'attribution (ou le changement incrémentiel) est une question d'évaluation importante, il est préférable de ne pas s'en tenir exclusivement à des modèles implicites, mais de les utiliser plutôt avec des éléments probants.

### 3.5 Utilisation des modèles de causalité pour l'évaluation

À la section 2.2 du présent chapitre, nous avons insisté sur la nature conceptuelle du modèle d'évaluation idéal ou classique. Dans ce modèle, la cause possible du résultat d'un programme est isolée grâce à l'utilisation de deux groupes en tout point identiques, exception faite de leur exposition au programme. À partir d'un modèle idéal, nous avons décrit d'autres modèles pouvant servir à attribuer des résultats à un programme, en précisant les divers degrés qui permettront à l'évaluateur de procéder par inférence et d'établir les obstacles à la validité interne correspondant à chacun d'entre eux.

Or, il existe une autre façon d'envisager les questions d'inférence causale, en utilisant un *modèle de causalité*, ce qui consiste à décrire l'influence marginale sur une *variable dépendante* d'une série de *variables indépendantes* choisies. Alors que le modèle quasi expérimental est axé sur des comparaisons entre les bénéficiaires du programme et les membres d'un ou plusieurs groupes témoins, le modèle de causalité se concentre sur les variables à inclure, tant endogènes (intrinsèques au programme) qu'exogènes (extérieures au programme), et sur les rapports de causalité postulés. Dans le modèle quasi expérimental, le programme est l'élément le plus important; dans le modèle de causalité, il ne constitue qu'une *variable indépendante* parmi d'autres, toutes censées influencer sur la *variable dépendante*.

Si nous revenons à notre exemple de l'évaluation d'un programme d'aide à un secteur d'activité industrielle dans lequel on compare les ventes à l'exportation réalisées par les entreprises qui bénéficient du programme à celles d'autres entreprises, un modèle de causalité tiendrait compte de variables telles que le secteur d'activité dans lequel les entreprises œuvrent, leur taille et le fait qu'elles ont bénéficié ou non du programme. À partir de là, l'évaluateur ferait une analyse de régression pour déterminer l'influence marginale de chacune de ces variables sur les ventes à l'exportation des entreprises intéressées.

De même, l'évaluation d'un programme de subventions à des organismes culturels dans diverses collectivités pourrait comparer a) les changements de l'assistance aux activités culturelles dans les collectivités qui bénéficient d'une importante subvention par tête et b) les changements de l'assistance dans les collectivités qui ont reçu une subvention moins importante. On pourrait générer un modèle de causalité des répercussions sur les niveaux d'assistance actuels du profil socio-économique et de l'infrastructure culturelle de la collectivité, ainsi que de ses tendances historiques à l'assistance à des activités culturelles. Les données ainsi obtenues pourraient remplacer à l'approche comparative que nous avons traitée ou s'y ajouter.

Dans la pratique, la plupart des évaluateurs préfèrent utiliser les deux méthodes à la fois pour déterminer les résultats d'un programme. Ils peuvent se servir d'un modèle quasi expérimental pour créer et manipuler des groupes témoins et, à partir de là, pour faire des inférences causales sur les résultats du programme, et peuvent aussi faire appel à un modèle de causalité pour obtenir une estimation de l'effet marginal des variables qui influent sur le succès du programme. Les recherches de Bickman

(1987) et Trochim (1986) ont produit des indications utiles sur la meilleure façon d'utiliser des modèles de causalité pour faire des évaluations.

Ces modèles sont particulièrement utiles dans les cas où une expérience empirique suffisante a confirmé avant l'évaluation l'existence de rapports entre les variables étudiées. En l'absence de modèle *a priori*, l'évaluateur aurait intérêt à utiliser la technique de l'appariement (constitution de blocs), comme nous l'avons vu aux sections 3.2.2 et 3.3.2, afin de recueillir des données pour les variables jugées importantes. En outre, il pourrait faire des analyses statistiques pour tenir compte des biais attribuables à la sélection ou aux événements historiques, afin d'accroître la validité de ces conclusions sur les effets du programme.

Les évaluateurs qui utilisent des modèles de causalité ont intérêt à consulter le chapitre 7 de l'ouvrage de Cook et Campbell (1979) intitulé *Quasi-experimentation*, qui contient une analyse des écueils à éviter lorsqu'on tente de faire des inférences causales fondées sur une «observation passive» (lorsqu'il n'y a pas formation de propos délibéré d'un groupe témoin). Deux des écueils les plus courants mentionnés dans cet ouvrage sont l'attention insuffisante accordée aux obstacles à la validité et l'utilisation de modèles structurels acceptables pour faire des prévisions, mais non des inférences causales.

### Références : Modèles de causalité

Bickman, L., éd., *Using Program Theory in Program Evaluation*, Vol. 33 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1987.

Blalock, H.M., Jr., éd., *Causal Models in the Social Sciences*, Chicago : Aldine, 1971.

Blalock, H.M., Jr., *Measurement in the Social Sciences: Theories and Strategies*, Chicago : Aldine, 1974.

Chen, H.T. et P.H. Rossi, «Evaluating with Sense: The Theory-Driven Approach», *Evaluation Review*, Vol. 7, 1983, p. 283 à 302.

Cook, T.D. et D.T. Campbell, *Quasi-experimentation*, Chicago : Rand-McNally, 1979, chapitres 4 et 7.

Cordray, D.S., «Quasi-experimental Analysis : A Mixture of Methods and Judgement», in W.M.K. Trochim, éd., *Advances in Quasi-experimental Design and Analysis*, Vol. 31 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1986, p. 9 à 27.

Duncan, B.D., *Introduction to Structural Equation Models*, New York : Academic Press, 1975.

Goldberger, A.S. et D.D. Duncan, *Structural Equation Models in the Social Sciences*, New York : Seminar Press, 1973.



Heise, D.R., *Causal Analysis*, New York : Wiley, 1975.

Mark, M.M., «Validity Typologies and the Logic and Practice of Quasi-experimentation», in W.M.K. Trochim, éd., *Advances in Quasi-experimental Design and Analysis*, Vol. 31 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1986, p. 47 à 66.

Rindskopf, D., «New Developments in Selection Modeling for Quasi-experimentation», in W.M.K. Trochim, éd., *Advances in Quasi-experimental Design and Analysis*, Vol. 31 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1986, p. 79 à 89.

Simon, H., «Causation», in D.L. Sill, éd., *International Encyclopedia of the Social Sciences*, Vol. 2, New York : Macmillan, 1968, p. 350 à 355.

Stolzenberg, J.R.M. et K.C. Land, «Causal Modeling and Survey Research», in Rossi, P.H., *et al.*, éd., TITRE MANQUANT, Orlando : Academic Press, 1983, p. 613 à 675.

Trochim, W.M.K., éd., *Advances in Quasi-experimental Design and Analysis*, Vol. 31 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1986.

### 3.6 Résumé

Le choix du modèle d'évaluation optimal est une tâche difficile, et c'est aussi l'aspect le plus important de la sélection d'une stratégie d'évaluation, puisque l'exactitude de l'information et des données — donc des preuves — produites dans ce contexte est largement fonction de la rigueur du modèle utilisé. C'est pour cette raison que l'évaluateur devrait s'efforcer d'opter pour le modèle le plus rigoureux possible compte tenu du temps et des ressources dont il dispose, ainsi que des autres facteurs d'ordre pratique. Le modèle choisi devrait être le plus près possible du modèle idéal (modèle expérimental). La rigueur du modèle d'évaluation et la crédibilité des résultats faiblissent à mesure que l'évaluateur descend la barre en passant d'un modèle expérimental à un modèle quasi expérimental, puis à un modèle implicite. Quel que soit le modèle choisi, il serait souhaitable d'inclure des éléments du modèle de causalité, pour renforcer la crédibilité des constatations.

Il arrive souvent que le seul modèle utilisable soit relativement peu rigoureux. Dans ce cas, l'évaluateur devrait cerner explicitement les principaux obstacles à la crédibilité des conclusions qu'il tire, afin de nuancer ses constatations en conséquence. Il devrait aussi rechercher d'autres modèles d'évaluation en vue de les utiliser pour étayer ses conclusions et pour atténuer les obstacles à la validité de sa démarche, voire pour ces deux raisons à la fois.

Bref, l'évaluateur devrait préciser explicitement le genre de modèle d'évaluation auquel il a recours pour chaque stratégie d'évaluation.

Parfois, on fait une évaluation sans bien comprendre le modèle utilisé, ce qui sape la crédibilité de l'information et des données obtenues, puisqu'on ne saisit pas bien le fondement de la «preuve». En précisant explicitement le modèle, l'évaluateur peut analyser ouvertement les principaux obstacles et trouver les arguments logiques ou d'autres éléments d'information susceptibles de les éliminer, de les atténuer ou d'en tenir compte de façon à renforcer la crédibilité globale de son évaluation.

Pour chaque modèle de recherche utilisé, l'évaluateur devrait dresser la liste des principaux obstacles plausibles à la validité, en analysant les implications de chacun.

Les auteurs ne s'entendent pas sur les obstacles à la validité que tel ou tel modèle permet généralement d'éliminer. Cronbach (1982), par exemple, conteste un grand nombre d'affirmations sur les obstacles à la validité décrits dans l'ouvrage plus classique de Cook et Campbell (1979). Ce désaccord est toutefois moins fréquent à l'égard d'évaluations données et de leurs modèles. En effet, dans chaque cas, il est habituellement évident s'il existe ou non d'autres explications plausibles d'un changement observé.

## Chapitre 4

# MÉTHODES DE COLLECTE DES DONNÉES

### 4.1 Introduction

Sans données pertinentes, il est impossible d'établir des rapports entre un programme et ses résultats. En outre, les méthodes de collecte doivent être choisies en fonction de la nature des données nécessaires et des sources accessibles. La nature des données dépendra elle-même de la méthode d'évaluation adoptée, des indicateurs utilisés pour obtenir les résultats des programmes et du type d'analyse à faire.

Il existe plusieurs façons de classer des données. Par exemple, on fait souvent une distinction entre les données *quantitatives* et les données *qualitatives*. Les **données quantitatives** sont des observations numériques, alors que les **données qualitatives** sont des observations correspondant à des catégories (p. ex., pour la couleur, rouge ou bleu, ou pour le sexe des participants, homme ou femme).

On fait aussi une distinction entre les données *subjectives* et les données *objectives*. Les **données subjectives** sont indissociables des sentiments, des attitudes et des perceptions personnelles, tandis que les **données objectives** sont fondées sur des *faits* observables qui — en théorie du moins — ne font pas appel au jugement personnel. Cela dit, les données subjectives et objectives peuvent toutes être mesurées d'une façon quantitative ou qualitative.

Il est possible aussi de faire une autre distinction, entre les données *longitudinales* et les données *transversales*. Les **données longitudinales** sont recueillies sur une certaine période, tandis que les **données transversales** le sont simultanément auprès de différentes entités, telles que des provinces ou des écoles.

Enfin, les données peuvent être classées en fonction de leur source : les **données primaires** sont recueillies par l'évaluateur à la source même; les **données secondaires**, elles, sont recueillies et consignées par une autre personne ou une autre organisation, parfois à des fins différentes de celles de l'évaluation.

Dans ce chapitre, nous avons examiné six méthodes de collecte des données utilisées pour l'évaluation d'un programme : le *dépouillement de la documentation spécialisée*, l'*étude de dossiers*, les *observations directes (sur le terrain)*, les *enquêtes*, la *consultation de spécialistes* et les *études de cas*. Les deux premières méthodes servent à recueillir des données secondaires, alors que les quatre dernières sont employées pour réunir des données primaires. Quoi qu'il en soit, chacune des six peut

être utilisée pour recueillir des données quantitatives et qualitatives. En outre, chacune pourrait être utilisée avec chacun des modèles exposés au chapitre précédent. Cependant, certaines méthodes de collecte de données se prêtent mieux à des modèles donnés.

Il convient de souligner que, même si les méthodes de collecte de données examinées dans ce chapitre sont surtout considérées comme des éléments d'une stratégie de recherche, la collecte de données elle-même est aussi extrêmement utile pour d'autres aspects d'une évaluation. D'ailleurs, plusieurs techniques de collecte sont d'excellents instruments pour alimenter la réflexion initiale sur les stratégies d'évaluation elles-mêmes et pour faciliter d'autres aspects exploratoires des études d'évaluation. Une enquête pourrait, par exemple, aider à cerner les questions connexes à l'évaluation. Une étude des dossiers peut aider à préciser les sources de données disponibles ou celles les plus facilement accessibles.

### **Références : Méthodes de collecte des données**

Cook, T.D. et C.S. Reichardt, *Qualitative and Quantitative Methods in Evaluation Research*, Thousand Oaks : Sage Publications, 1979.

Delbecq, A.L., et al., *Group Techniques for Program Planning: A Guide to Nominal Group and Delphi Processes*, Glenview : Scott, Foresman, 1975.

Dexter, L.A., *Elite and Specialized Interviewing*, Evanston (Illinois) : Northwestern University Press, 1970.

Gauthier, B., éd., *Recherche sociale : de la problématique à la collecte des données*, Montréal : Les Presses de l'Université du Québec, 1984.

Kidder, L.H. et M. Fine, «Qualitative and Quantitative Methods: When Stories Converge», in *Multiple Methods in Program Evaluation*, Vol. 35 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1987.

Levine, M., «Investigative Reporting as a Research Method: An Analysis of Bernstein and Woodward's *All The President's Men*», *American Psychologist*, Vol. 35, 1980, p. 626 à 638.

Miles, M.B. et A.M. Huberman, *Qualitative Data Analysis: A Sourcebook and New Methods*, Thousand Oaks : Sage Publications, 1984.

Patton, M.Q., *Qualitative Evaluation Methods*, Thousand Oaks : Sage Publications, 1980.

Martin, Michael O. et V.S. Mullis, éd., *Quality Assurance in Data Collection*, Chestnut Hill : Center for the Study of Testing, Evaluation, and Educational Policy, Boston College, 1996.

Stouthamer-Loeber, Magda et Bok van Kammen, Welmoet, *Data Collection and Management: A Practical Guide*, Thousand Oaks : Sage Publications, 1995.

Webb, E.J., *et al.*, *Nonreactive Measures in the Social Sciences* (2<sup>e</sup> édition), Boston : Houghton Mifflin, 1981.

Weisberg, Herbert F., Krosnick, Jon A. et Bruce D. Bowen, éd., *An Introduction to Survey Research, Polling, and Data Analysis*, Thousand Oaks : Sage Publications, 1996.

**Notes :**

## 4.2 Dépouillement de la documentation spécialisée

L'évaluateur dépouille la documentation pour prendre connaissance des travaux effectués dans le domaine à l'étude et donc pour profiter des expériences, des constatations et des erreurs de prédécesseurs ayant fait des recherches analogues ou connexes aux siennes. C'est un exercice qui peut lui fournir des indications d'une valeur inestimable sur le secteur de programme visé, et c'est pourquoi il doit toujours se faire dès le début de l'étude d'évaluation.

Le dépouillement porte sur deux types d'ouvrages et de documents, d'abord les documents officiels, les rapports de recherche généraux, les articles publiés et les livres portant sur le secteur du programme, qui permettent à l'évaluateur de se familiariser avec les théories et les concepts relatifs au programme et de s'informer des généralisations susceptibles de s'appliquer aux questions qui l'intéressent. Il peut aussi y trouver d'autres points d'évaluation et des méthodes auxquelles il n'avait pas pensé, ce qui peut l'aider à réaliser une évaluation plus efficace. Par exemple, des recherches antérieures sur les programmes d'aide à un secteur d'activité industrielle donné peuvent laisser entendre que l'efficacité des interventions varie énormément selon la taille des entreprises, ce qui signifie que toute méthode d'échantillonnage utilisée pour l'évaluation doit assurer une représentation suffisante d'entreprises de toutes les tailles (grâce à un échantillonnage aléatoire par blocs) afin que les résultats de l'évaluation puissent être généralisés.

L'évaluateur passe ensuite en revue les études spéciales (y compris les évaluations antérieures) sur le secteur qui l'intéresse. Dans ce contexte, il peut aller jusqu'à compiler et résumer les constatations de ces études et se servir de ces renseignements comme d'intrants pour diverses composantes de son évaluation. Ainsi, l'évaluateur chargé d'étudier un programme d'aide à un secteur d'activité industrielle donné pourrait trouver dans des rapports d'études antérieures des données sur l'emploi dans les régions où l'importance de l'aide reçue a varié énormément. Il peut alors avoir recours à un modèle quasi expérimental pour intégrer ces renseignements à l'évaluation, les régions ayant reçu beaucoup d'aide constituant un groupe et celles qui en ont reçu moins servant de groupe témoin.

### Avantages et inconvénients

Le dépouillement de la documentation spécialisée dès le début de l'évaluation peut permettre à l'évaluateur de gagner du temps, d'économiser de l'argent et de ménager ses efforts. En effet, un dépouillement minutieux a généralement plusieurs avantages.

Les recherches antérieures peuvent laisser entrevoir des hypothèses à vérifier ou des questions à examiner dans le cadre de l'étude.

Par exemple, au chapitre 3, nous avons insisté sur l'importance d'identifier le plus tôt possible les explications concurrentes d'un résultat observé, autres que celles de l'intervention du programme. L'analyse des recherches antérieures peut révéler certaines possibilités d'explications concurrentes (autrement dit des obstacles à la validité). En pareil cas, il faudrait opter pour une stratégie d'évaluation permettant d'isoler l'effet du programme de ces autres explications.

Le dépouillement de la documentation peut mettre en lumière des difficultés méthodologiques particulières et indiquer à l'évaluateur des techniques et des moyens de les contourner.

Dans certains cas, les recherches antérieures fournissent directement des réponses aux questions d'évaluation, ce qui permet d'éviter un travail de collecte de données inutile.

On peut aussi trouver des sources de données secondaires utilisables dans les études antérieures, ce qui réduit le besoin de recueillir des données primaires.

Même lorsque les données secondaires ne répondent pas directement aux questions posées dans l'évaluation, elles pourraient être utilisées de pair avec les données primaires, comme intrants dans la stratégie d'évaluation ou encore comme données de référence, pour vérifier la validité.

**Le dépouillement de la documentation spécialisée est un moyen relativement économique et efficace de recueillir des données pertinentes, et il est souvent rentable. Il doit toujours se faire au cours de l'étude préparatoire à l'évaluation.** C'est aussi un exercice très utile pour trouver de nouvelles hypothèses et cerner d'éventuelles difficultés méthodologiques, ainsi que pour tirer et étayer des conclusions. Les renseignements qu'on y glane peuvent servir d'intrants pour d'autres techniques de collecte des données.

Les inconvénients des données obtenues en dépouillant la documentation sont inhérents à la nature même de la plupart des données secondaires, qui ont été produites pour une autre fin que celle de l'évaluation à réaliser.

Il est possible que les données et l'information recueillies grâce au dépouillement de la documentation ne soient pas suffisamment pertinentes ou compatibles avec l'objet de l'évaluation pour pouvoir être utilisées dans l'étude.

Les données secondaires sont dites pertinentes dans la mesure où elles correspondent à l'objet de l'évaluation, ce qui signifie qu'elles doivent être compatibles avec ses exigences. Par exemple, des données secondaires à l'échelle nationale ne seraient guère utiles pour une évaluation fondée sur des données par province. En outre, les échelles de mesure doivent être compatibles, elles aussi. Quand l'évaluateur a besoin de données sur les enfants de 8 à 12 ans, des données

secondaires sur les 5 à 9 ans ou les 10 à 14 ans ne lui servent pas à grand-chose. Enfin, comme le facteur temps influe largement sur la pertinence des données, les données secondaires sont bien souvent tout simplement trop vieilles pour être utilisables. (N'oublions pas qu'il s'écoule ordinairement de un à trois ans entre la collecte des données et la publication.)

Il est souvent difficile de déterminer l'exactitude des données secondaires.

En fait, c'est vraiment l'aspect névralgique des données secondaires. L'évaluateur n'a bien sûr rien eu à dire sur la méthode utilisée pour les recueillir, mais il doit quand même en évaluer la validité et la fiabilité. C'est pour cette raison qu'il lui est fortement recommandé de se fonder chaque fois qu'il le peut sur la source initiale des données secondaires, c'est-à-dire le rapport initial, étant donné que ce document est généralement plus complet qu'un deuxième ou un troisième document faisant appel aux données en question et qu'il contient souvent des avertissements, des restrictions et des détails d'ordre méthodologique qui ne sont pas signalés dans les documents qui le citent.

Bref, le dépouillement exhaustif de la documentation spécialisée est un moyen rapide et relativement peu coûteux d'obtenir de l'information conceptuelle et empirique sur le contexte d'une évaluation. Il s'ensuit que l'évaluateur devrait s'y astreindre dès le début de son étude. Néanmoins, il devrait évaluer soigneusement la pertinence et l'exactitude des données qui s'y trouvent, dans toute la mesure du possible. Enfin, il devrait se méfier, en s'efforçant de ne pas accorder trop d'importance à des données secondaires quand il a très peu d'information sur les méthodes utilisées pour les recueillir.

### **Références : Dépouillement de la documentation spécialisée**

Goode, W.J. et Paul K. Hutt, *Methods in Social Research*, New York : McGraw-Hill, 1952, chapitre 9.

Katz, W.A., *Introduction to Reference Work: Reference Services and Reference Processes, Volume II*, New York : McGraw-Hill, 1982, chapitre 4.

### **Notes :**



### 4.3 Étude de dossiers

Tout comme le dépouillement de la documentation spécialisée, l'étude de dossiers est une méthode de collecte des données conçue afin de faciliter l'obtention de données utilisables dans l'évaluation. C'est toutefois une démarche qui se prête mieux que la première à une familiarisation avec le programme faisant l'objet de l'évaluation. Les données déjà recueillies sur lui — et sur ses résultats — peuvent réduire les besoins de collecte de nouvelles données, sensiblement comme on l'a vu pour le dépouillement de la documentation.

Il existe habituellement deux types de dossiers, les dossiers généraux portant sur un programme et les dossiers relatifs à des projets, des clients ou des participants donnés. Le type de dossier que les gestionnaires de programmes conservent dépend du programme lui-même. Par exemple, dans le cas d'un programme d'aide financière à des projets d'économie d'énergie, on pourrait avoir des dossiers sur chaque projet, chaque client (auteur du projet) et chaque participant (collaborateur au projet). Par contre, dans un programme de formation destiné à des spécialistes en soins de santé dans les collectivités du Nord, on pourrait ne conserver des dossiers que sur les spécialistes qui ont participé aux séances de formation. Dans la pratique, il y a un type d'examen pour chacun de ces types de dossiers : un examen général portant sur les dossiers d'un programme et un examen plus systématique des dossiers relatifs à chaque projet, chaque client ou chaque participant.

Les études de dossiers peuvent porter sur les types suivants de documents relatifs aux programmes :

- documents du Cabinet, documents relatifs à la négociation et à la mise en œuvre d'un protocole d'entente avec le Conseil du Trésor, présentations au Conseil du Trésor, plans d'activités ou rapports sur le rendement d'un ministère, rapports du vérificateur général et procès-verbaux des réunions du comité exécutif d'un ministère;
- dossiers administratifs, portant notamment sur la taille du programme ou du projet, sur le type de participants et sur leur expérience, sur l'expérience postérieure au projet, sur les coûts du programme ou du projet et sur les mesures des caractéristiques des participants avant et après la mise en œuvre du programme;
- dossiers sur les participants, comprenant notamment des données socio-économiques (âge, sexe, région, revenu, profession, etc.), des dates critiques (admission au programme), des données de suivi et la description des événements importants (changements d'emploi, déménagements, etc.);

- dossiers sur les projets et les programmes, portant notamment sur les événements critiques (lancement des projets et rencontres avec les cadres supérieurs, par exemple), le personnel des projets (roulement du personnel, etc.) et les événements ainsi que les modifications qui ont marqué la mise en œuvre des projets;
- dossiers financiers.

Les données figurant dans les dossiers peuvent être conservées dans le système informatisé de gestion de l'information du programme, ou encore sur papier. Elles peuvent avoir été recueillies expressément pour fins d'évaluation, s'il y a eu une entente préalable au sujet d'un cadre d'évaluation.

### Avantages et inconvénients

Les études de dossiers peuvent être utiles à trois titres au moins.

1. L'étude des dossiers généraux du programme peut fournir des données contextuelles et des renseignements d'une valeur inestimable sur le programme et sur son milieu permettant de bien situer les résultats du programme dans leur contexte.

Ce genre d'étude peut générer des renseignements contextuels fondamentaux sur le programme (mandat, historique, politiques, style de gestion et contrainte, par exemple) grâce auxquels l'évaluateur peut se familiariser avec lui. En outre, elle peut produire des renseignements clés pour les spécialistes de l'extérieur dans le secteur d'un programme (voir la section 4.6) et générer des intrants pour une éventuelle analyse qualitative (voir la section 5.4).

2. L'étude des dossiers de personnes ou de projets peut fournir des indicateurs sur les résultats du programme.

Par exemple, dans une étude portant sur un programme d'aide internationale, un examen des dossiers de projets pourrait fournir des mesures des résultats telles que le rapport produit-capita, valeur ajoutée-unité de capital, productivité du capital employé, intensité de capital, emploi-unité de capital, valeur ajoutée-unité d'intrant total et diverses autres fonctions de production. Si ces mesures ne permettent pas d'évaluer directement l'efficacité du programme, elles constituent néanmoins des indicateurs susceptibles de servir d'intrants pour l'évaluation. Enfin, les données ainsi obtenues peuvent se révéler suffisantes pour qu'on puisse procéder à une analyse coûts-avantages ou coût-efficacité (voir la section 5.6).

3. L'étude de dossiers peut fournir un cadre et une base utiles pour la collecte d'autres données.

L'étude de dossiers peut notamment permettre de déterminer la population (base d'échantillonnage) de laquelle l'échantillon d'enquête doit être tiré. Les renseignements contextuels tirés des dossiers peuvent servir à créer l'échantillon le plus puissant possible et à préparer l'enquêteur à réaliser ses entrevues. On rebute toujours les gens en leur demandant des renseignements qui se trouvent déjà dans les dossiers, et c'est pourquoi il faudrait recueillir toute l'information qui s'y trouve avant de commencer l'enquête.

L'étude de dossiers a des avantages certains du fait même qu'elle est éminemment faisable.

L'étude de dossiers peut être relativement peu coûteuse.

En fait, ce genre d'étude a fort peu de retombées gênantes pour les personnes et les groupes qui ne sont pas visés par l'administration du programme. Tout comme le dépouillement de la documentation spécialisée, l'étude de dossiers est pour l'évaluateur un moyen aussi fondamental que naturel de se familiariser avec le programme. En outre, c'est une excellente façon de lui permettre d'éviter une collecte coûteuse de nouvelles données lorsqu'il existe déjà des données pertinentes.

Toutefois, l'étude de dossiers présente aussi certains inconvénients.

Les dossiers relatifs aux programmes sont souvent incomplets, ou inutilisables pour d'autres raisons.

Plus souvent qu'autrement, on n'accorde qu'une importance secondaire au système de classement central, qui ne renferme alors que de courtes notes de service des comités, des comptes rendus des décisions finales, ainsi de suite. En rétrospective, ces dossiers ne brossent qu'un tableau incomplet.

Dans sa recherche des documents qui ont inspiré une politique, un programme ou un projet, l'évaluateur peut constater que l'information figure dans des dossiers détenus par différentes personnes plutôt que dans un dépôt central des dossiers du programme. Cela peut causer plusieurs difficultés. Par exemple, l'expérience semble laisser entendre que, lorsque le projet progresse au-delà de l'exécution du mandat du groupe de travail, les participants ferment leurs dossiers plutôt que de les tenir à jour. De même, lorsque quelqu'un cesse de participer à un groupe de travail, ses dossiers sont souvent perdus, et, comme le rôle des participants change rapidement dans les premières étapes d'un programme, il devient très difficile d'en trouver un qui soit exhaustif.

L'étude de dossiers fournit rarement des renseignements sur les groupes témoins, sauf dans des circonstances exceptionnelles, par exemple lorsqu'il existe des dossiers sur les candidats au programme qui ont été refusés.

Pour évaluer efficacement les répercussions du programme, l'évaluateur doit avoir accès à un groupe témoin quelconque. Dans le contexte de l'étude de dossiers, cela suppose qu'il doit obtenir des renseignements sur les participants avant qu'ils ne bénéficient du programme, ou encore des renseignements sur les non-participants. Malheureusement, ces renseignements existent rarement, sauf si un cadre d'évaluation a été approuvé et mis en œuvre au préalable. Faute de données de ce genre, l'évaluateur peut se voir contraint de recueillir de nouvelles données qui risquent de ne pas être comparables avec celles des dossiers originaux.

Il reste toutefois que l'étude de dossiers peut fournir des renseignements sur des groupes témoins lorsque les niveaux du programme varient; c'est utile lorsqu'on applique un modèle dont les mesures sont prises après le programme seulement avec traitement différent. L'examen des dossiers peut aussi fournir à l'évaluateur l'information de base dont il a besoin pour définir et choisir un groupe témoin.

En dépit des limites de cette méthode, il **faudrait faire une étude de dossiers** à l'étape de l'étude préparatoire à l'évaluation afin de déterminer le type de données disponibles et leur pertinence pour les questions à évaluer. Cette démarche génère aussi les renseignements nécessaires à l'examen de questions d'évaluation particulières (par exemple des renseignements contextuels et des indicateurs éventuels des résultats du programme).

### **Références : Analyse des données secondaires**

Boruch, R.F., *et al.*, *Reanalyzing Program Evaluations – Policies and Practices for Secondary Analysis for Social and Education Programs*, San Francisco : Jossey-Bass, 1981.

Weisler, Carl E., U.S. General Accounting Office, *Review Topics in Evaluation: What Do You Mean by Secondary Analysis?*

### **Notes :**

## 4.4 Observations directes

Le vieux proverbe «Voir, c'est croire» tient toujours, et l'observation directe apporte généralement des preuves plus convaincantes que les sources secondaires. Aller sur le terrain obtenir des renseignements de première main sur le sujet de l'évaluation est un moyen très efficace de recueillir de l'information et des données. La consignation des résultats de ces observations en photos ou sur vidéo peut aussi être très utile, et les documents ainsi obtenus peuvent avoir un impact considérable sur le lecteur des rapports d'évaluation.

L'observation comporte la sélection, l'observation et la consignation des objets, des événements ou des activités qui jouent un rôle important dans l'administration du programme à évaluer. Les conditions observées peuvent ensuite être comparées à des critères préétablis, et les différences par rapport à ces critères peuvent être analysées pour en déterminer l'importance.

Dans certains cas, l'observation directe est un instrument essentiel pour comprendre le fonctionnement d'un programme. Par exemple, une équipe faisant l'évaluation du dédouanement dans les aéroports pourrait observer de longues files d'attente chaque fois que deux 747 atterriraient à la même heure. L'achalandage qui en résulterait réduirait à la fois l'efficacité de l'inspection et la qualité du service aux voyageurs. Prenons un autre exemple, celui d'un cas où des produits chimiques dangereux auraient été entreposés au mépris des règles de sécurité : il en aurait résulté des conditions de travail dangereuses pour le personnel et une violation de la réglementation en matière de santé et de sécurité au travail. Or, aucune de ces constatations n'aurait été remarquée par l'évaluateur s'il s'était contenté d'étudier des documents écrits.

Les données obtenues grâce à l'observation directe servent à décrire le contexte d'un programme, les activités qui s'y déroulent, les personnes qui y ont participé et la signification de ces activités pour elles. C'est une méthode qui a été largement utilisée par les spécialistes du comportement comme les anthropologues et les psychosociologues. Elle permet à l'évaluateur d'obtenir d'une façon holistique des données sur un programme et sur ses répercussions.

L'observation directe nécessite des visites aux endroits où le programme est exécuté afin de voir ce qui se passe et de prendre des notes. Les participants et le personnel peuvent savoir qu'on les observe ou l'ignorer.

Les rapports d'observation devraient être rédigés immédiatement après la visite sur place, avec suffisamment de détails descriptifs pour que le lecteur puisse comprendre ce qui s'est produit, et de quelle façon. Les descriptions doivent être factuelles, précises et complètes, mais sans détails superflus. Les données de ce genre sont utiles pour l'évaluation, parce que l'évaluateur et les utilisateurs peuvent comprendre les activités et les effets d'un programme grâce à une information descriptive détaillée sur ce qui s'est produit et sur la réaction des personnes intéressées.

## Avantages et inconvénients

L'observation ne fournit que des renseignements anecdotiques, à moins d'être combinée avec un programme planifié de collecte de données. En effet, un événement aléatoire ne peut servir de base de généralisation. Certains éléments d'observation directe peuvent être justifiés dans presque toutes les évaluations, mais la planification et la conduite des activités de collecte de données représentatives sur le terrain peut entraîner des coûts élevés.

L'information permet à l'évaluateur de mieux comprendre un programme, particulièrement si celui-ci met en jeu une technique ou un processus complexe ou perfectionné. Grâce à l'observation personnelle directe, l'évaluateur est en mesure de se faire une idée complète du fonctionnement du programme. En outre, c'est une approche qui lui permet d'aller au-delà des perceptions sélectives d'autres personnes qu'il obtient grâce à des entrevues, par exemple. S'il est lui-même un observateur sur le terrain, l'évaluateur aura ses propres perceptions sélectives, mais pourra quand même présenter une vue plus complète du programme en intégrant ses propres perceptions aux données dont il dispose.

L'observation directe permet à l'évaluateur de saisir des détails qui risquent d'échapper aux membres du personnel ou de constater des éléments qu'ils hésitent à soulever au cours d'une entrevue.

La plupart des organisations ont des activités répétitives que les participants considèrent comme tout à fait normales. Par conséquent, des particularités importantes peuvent leur échapper totalement, alors qu'elles sont facilement perçues par des non-participants. C'est pour cette raison qu'un étranger — en l'occurrence l'évaluateur — peut souvent avoir un point de vue «nouveau», et c'est pourquoi le contact direct avec la situation permet à l'évaluateur d'obtenir une information qu'il ne pourrait pas recueillir autrement.

La fiabilité et la validité des observations sont fonction de la compétence de l'observateur et de sa sensibilisation à ses biais.

L'observation directe ne peut jamais être répétée, puisque deux personnes qui observent les mêmes activités peuvent aboutir à des observations différentes. Il s'ensuit que la validité interne et la validité externe des données obtenues grâce aux observations directes sont limitées.

Le personnel du programme peut avoir un comportement très différent de la normale lorsqu'il sait qu'un évaluateur l'observe.

L'évaluateur doit être conscient de ce phénomène, en sachant que la personne, les participants ou les deux peuvent changer nettement de comportement lorsqu'ils se savent observés. Il doit prendre les mesures nécessaires pour éviter cet écueil, ou du moins pour tenir compte de son influence.

**Références : Observations directes**

Canada, Bureau du Vérificateur général du Canada, *Bulletin 84-7, Photographies et autres aides visuelles*. (Ce bulletin porte sur l'utilisation ultime des photographies dans le Rapport annuel, mais il contient aussi des explications des caractéristiques qui rendent une photographie efficace comme élément de preuve.)

Guba, E.G., «Naturalistic Evaluation», in Cordray, D.S., et al., éd., *Evaluation Practice in Review*, Vol. 34 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1987.

Guba, E.G. et Y.S. Lincoln, *Effective Evaluation: Improving the Usefulness of Evaluation Results through Responsive and Naturalistic Approaches*, San Francisco : Jossey-Bass, 1981.

Patton, M.Q., *Qualitative Evaluation Methods*, Thousand Oaks : Sage Publications, 1980.

Pearsol, J.A., éd., «Justifying Conclusions in Naturalistic Evaluations», *Evaluation and Program Planning*, Vol. 10, N° 4, 1987, p. 307 à 358.

V. Van Maasen, J., éd., *Qualitative Methodology*, Thousand Oaks : Sage Publications, 1983.

Webb, E.J., et al., *Nonreactive Measures in the Social Sciences* (2<sup>e</sup> édition), Boston : Houghton Mifflin, 1981.

Williams, D.D., éd., *Naturalistic Evaluation*, Vol. 30 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1986.

**Notes :**

## 4.5 Enquêtes

Dans le contexte d'une évaluation, les enquêtes sont des moyens systématiques de collecte de données primaires — quantitatives, qualitatives ou les deux — sur un programme et sur ses résultats auprès de personnes ou d'autres sources (comme des dossiers) qui y sont associés. On entend par «enquête» une démarche structurée conçue pour obtenir les données nécessaires d'un échantillon de la population visée (ou de toute cette population). La population visée est composée des personnes dont il faut obtenir des données et de l'information. Bien exécutée, l'enquête est un moyen précis et efficace pour déterminer les caractéristiques (physiques et psychologiques) d'à peu près n'importe quelle population étudiée.

On utilise souvent des enquêtes dans les évaluations, en raison de leur polyvalence. En fait, on peut s'en servir pour recueillir des données sur virtuellement n'importe quoi. Néanmoins, elles servent essentiellement à fournir des intrants pour une autre technique d'analyse, car elles ne sont pas des stratégies d'évaluation comme telles, mais simplement des méthodes de collecte de données.

La conception d'une enquête en vue d'une évaluation exige de la minutie et de la compétence. Il existe de nombreux guides sur la façon d'élaborer une enquête; certains sont mentionnés à la fin du présent chapitre. L'annexe I est une description et une analyse des éléments de base des enquêtes. Dans les pages qui suivent, nous décrivons succinctement la façon de mener une enquête dans le contexte d'une évaluation.

L'évaluateur devrait franchir les trois étapes suivantes avant de procéder à une enquête. Il lui faut d'abord déterminer l'information nécessaire à l'évaluation, puis mettre au point l'instrument propre à recueillir cette information et enfin le mettre à l'essai. Ces trois étapes s'appliquent à toutes les techniques de collecte des données. Nous les expliquons ici dans le contexte des enquêtes, parce qu'elles sont très courantes dans les évaluations.

### a) Déterminer l'information nécessaire à l'évaluation

La première étape — qui est aussi la plus fondamentale — consiste à déterminer aussi précisément que possible l'information nécessaire à l'examen d'une question d'évaluation donnée.

Premièrement, l'évaluateur doit bien comprendre la question à l'étude, afin de déterminer quelle sorte de données ou d'information lui donnera des indications utiles. Il doit aussi se demander quoi faire de l'information une fois qu'il l'aura recueillie. Quels genres de tableaux produira-t-il? Quelles sortes de conclusions souhaite-t-il tirer? Si le travail n'est pas fait avec soin à cette étape, l'évaluateur risque soit de recueillir trop d'information, soit de finir par constater que des éléments clés lui font défaut.



Deuxièmement, l'évaluateur doit s'assurer que les données dont il a besoin sont introuvables ailleurs ou ne peuvent pas être recueillies de façon plus efficace et plus pertinente avec d'autres méthodes de collecte de données. Dans n'importe quel secteur de programme, il se peut qu'on ait déjà effectué des enquêtes ou qu'il y ait des enquêtes en cours. L'évaluateur doit donc absolument dépouiller la documentation pour déterminer si les données dont il a besoin n'existent pas déjà ailleurs.

Enfin, l'évaluateur doit tenir compte des impératifs d'économie et d'efficacité. En effet, il est toujours tentant de recueillir des renseignements qu'il serait bon d'avoir, mais qui ne sont pas indispensables. L'évaluateur devrait savoir que le coût de l'enquête est largement fonction de son envergure et de sa nature, et que la collecte de données «supplémentaires» fait inévitablement augmenter les coûts.

### **b) Mettre au point l'instrument de collecte de l'information**

L'annexe 1, intitulée «Enquêtes» porte notamment sur la préparation de l'enquête même. Dans ce contexte, il s'agit de déterminer l'échantillon à utiliser, de choisir la méthode d'enquête la mieux appropriée et de concevoir le questionnaire. Ces étapes, plutôt itératives que successives, sont déterminées par les besoins d'information à mesure qu'on les constate.

### **c) Mettre l'instrument d'enquête à l'essai**

Au moment de leur utilisation sur le terrain, on constate souvent de graves lacunes dans les instruments d'enquête qui n'ont pas été mis à l'essai comme il se doit. Il faut absolument faire un essai préliminaire auprès d'un échantillon représentatif de la population visée, afin de valider aussi bien le questionnaire que les méthodes que l'on souhaite utiliser pour mener l'enquête. Cet essai fournira des renseignements sur les éléments suivants :

- *Clarté des questions*

Le libellé des questions est-il assez clair? Les répondants interprètent-ils tous les questions de la même façon? L'enchaînement des questions est-il logique?

- *Taux de réponse*

Les répondants trouvent-ils certaines questions dérangeantes? La technique d'entrevue les agace-t-elle? Refusent-ils de répondre à certaines parties du questionnaire?

- *Longueur et durée*

Combien de temps faut-il pour répondre au questionnaire?

- *Méthode d'enquête*

S'il s'agit d'une enquête postale, le taux de réponse est-il satisfaisant? Existe-t-il une autre méthode susceptible d'obtenir le taux de réponse recherché?

### **Avantages et inconvénients**

Les avantages et les inconvénients des méthodes d'enquête sont étudiés à la section A.5 de l'annexe 1, mais voici déjà quelques observations d'ordre général.

Une enquête est un moyen de recueillir des données auprès d'une population.

Avec une enquête, il est possible d'obtenir des données attitudinales sur virtuellement n'importe quel aspect d'un programme et sur ses résultats. La population visée peut être importante ou réduite, et l'enquête peut comprendre une série temporelle de mesures ou des mesures prises auprès de populations variées.

Une enquête bien menée produit des renseignements fiables et valides.

Il existe de nombreuses techniques d'enquête raffinées. En outre, on peut lire bien des livres, suivre des cours, consulter des spécialistes et des entreprises d'experts-conseils du secteur privé pour veiller à ce que l'information recueillie soit pertinente, opportune, valide et fiable.

Néanmoins, considérées comme instruments de collecte de données, les enquêtes présentent plusieurs inconvénients.

La conception, l'exécution et l'interprétation des enquêtes exigent de la compétence. Il est facile de les utiliser à mauvais escient, ce qui produit des données et de l'information non valides.

De nombreux facteurs peuvent fausser les méthodes d'enquête et compromettre la fiabilité et la validité des données recueillies : un échantillonnage biaisé, un biais de non-réponse, la sensibilité des répondants au questionnaire, un biais attribuable à l'enquêteur et des erreurs de codage. Il faut tenir compte de chacun de ces écueils éventuels. Statistique Canada a préparé un répertoire de méthodes d'évaluation de la qualité des données recueillies dans les enquêtes (1982).

La qualité des enquêtes doit faire l'objet d'un contrôle rigoureux. Or, il arrive souvent que l'évaluateur confie l'étape de la collecte des données à des contractuels. Dans ces cas-là, il est sage que la fiabilité du travail de l'entrepreneur soit vérifiée, notamment par des entrevues de contrôle auprès d'un petit échantillon de répondants.

**Références : Enquêtes**

- Babbie, E.R., *Survey Research Methods*, Belmont : Wadsworth, 1973.
- Bradburn, N.M. et S. Sudman, *Improving Interview Methods and Questionnaire Design*, San Francisco : Jossey-Bass, 1979.
- Braverman, Mark T. et Jana Kay Slater, *Advances in Survey Research*, Vol. 70 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1996.
- Canada, Secrétariat du Conseil du Trésor, *Mesure de la satisfaction des clients : Concevoir et adopter de saines pratiques de mesure et de suivi de la satisfaction des clients*, Ottawa, octobre 1991.
- Canada, Statistique Canada, *Lignes directrices concernant la qualité* (2<sup>e</sup> édition), Ottawa, 1987.
- Canada, Statistique Canada, *Répertoire des méthodes d'évaluation des erreurs dans les recensements et les enquêtes*, Ottawa, 1982, CSCCB-F.
- Dexter, L.A., *Elite and Specialized Interviewing*, Evanston (Illinois) : Northwestern University Press, 1970.
- Fowler, Vol. et J. Floyd, *Improving Survey Questions: Design and Evaluation*, Thousand Oaks : Sage Publications, 1995.
- Gliksman, Louis, *et al.*, «Responders vs. Non-Responders to a Mail Survey: Are They Different?», *Canadian Journal of Program Evaluation*, Vol. 7, N<sup>o</sup> 2, octobre-novembre 1992, p. 131 à 138.
- Kish, L., *Survey Sampling*, New York : Wiley, 1965.
- Robinson, J.P. et P.R. Shaver, *Measurement of Social Psychological Attitudes*, Ann Arbor: Survey Research Center, University of Michigan, 1973.
- Rossi, P.H., Wright, J.D. et A.B. Anderson, éd., *Handbook of Survey Research*, Orlando : Academic Press, 1985.
- Warwick, D.P. et C.A. Lininger, *The Survey Sample: Theory and Practice*, New York : McGraw-Hill, 1975.

## 4.6 Consultation de spécialistes

Cette méthode de collecte de données met à profit les perceptions et les connaissances des spécialistes de divers domaines fonctionnels, en tant qu'indicateurs d'évaluation. Fondamentalement, elle consiste à obtenir l'opinion de ces spécialistes sur des questions d'évaluation données. L'évaluateur utilise ensuite ces renseignements pour déterminer les résultats du programme. En fait, la consultation de spécialistes est un type d'enquête particulier, ce qui signifie que tout ce qui a été dit dans la section sur les enquêtes s'applique également ici. Toutefois, en raison de sa fréquence, la consultation de spécialistes doit faire l'objet d'une section distincte.

Il importe de préciser d'emblée que cette consultation est une méthode qui devrait en théorie servir à compléter (ou à remplacer, en l'absence d'indicateurs plus objectifs) d'autres mesures des résultats d'un programme. Rappelons-le, c'est une méthode de collecte de données qui ne consiste pas à joindre des spécialistes à l'équipe d'évaluation, mais plutôt à se servir d'eux comme sources de données pour étudier les questions à évaluer.

On peut recueillir des opinions de spécialistes et les résumer de façon systématique, bien que les résultats de cette démarche restent toujours subjectifs. Supposons par exemple que l'évaluateur cherche à déterminer de quelle façon un programme d'aide donnée a favorisé l'avancement des connaissances scientifiques. L'une des façons de mesurer ces deux variables difficiles à quantifier pourrait consister à interroger des spécialistes du domaine scientifique en question. L'évaluateur aurait recours à diverses méthodes — par exemple à une enquête postale ou à des entrevues individuelles — pour obtenir des mesures quantitatives. Dans ce contexte, il pourrait faire une enquête ponctuelle ou utiliser une technique interactive comme la méthode Delphi (voir Linstone et Turoff, 1995) ou encore la rétroaction qualitative contrôlée (voir Press, 1978).

### Avantages et inconvénients

La consultation de spécialistes peut servir à obtenir des mesures dans des domaines où l'on manque de données subjectives. C'est une technique de collecte de données rapide et relativement peu coûteuse.

Grâce à sa souplesse et à sa facilité d'utilisation, la consultation de spécialistes se prête à l'évaluation de presque n'importe quel résultat, voire de n'importe quel aspect d'un programme. Sa crédibilité est d'autant plus grande qu'elle est exécutée aussi systématiquement que possible. Néanmoins, elle présente plusieurs inconvénients importants.

L'évaluateur peut avoir de la difficulté à trouver suffisamment de spécialistes compétents pour assurer la fiabilité statistique des résultats.

Il risque d'être difficile d'amener les parties intéressées à s'entendre sur le choix des spécialistes.

Comme il est peu probable que les spécialistes soient tous aussi versés dans un domaine donné, il faudrait pondérer les résultats.

On peut bien sûr utiliser des méthodes statistiques pour tenter de pondérer la compétence inégale des spécialistes, mais ces méthodes manquent largement de précision, et c'est pourquoi l'évaluateur risque de considérer toute réponse comme étant d'égale importance.

La validité de la mesure peut être contestée, comme dans toutes les évaluations verbales.

Les spécialistes peuvent se fonder sur des critères différents ou attribuer une valeur différente aux chiffres sur les échelles de notation. Par exemple, le spécialiste qui évalue à 3 la contribution d'un projet à l'avancement des connaissances scientifiques, sur une échelle de 1 à 5, peut lui accorder la même valeur qu'un autre qui l'évalue à 4, la seule différence étant que les deux accordent une valeur différente au degré de l'échelle.

La crédibilité des spécialistes peut toujours être mise en doute, comme dans toute évaluation suggestive.

Les différences sur le choix et la valeur des spécialistes peuvent facilement faire perdre toute valeur même à un ensemble remarquable d'opinions.

Par conséquent, il faudrait éviter de faire de la consultation de spécialistes la seule source de données d'une évaluation.

## Références : Consultation de spécialistes

Boberg, Alice L. et Sheryl A. Morris-Khoo, «The Delphi Method: A Review of Methodology and an Application in the Evaluation of a Higher Education Program», *Canadian Journal of Program Evaluation*, Vol. 7, N° 1, avril-mai 1992, p. 27 à 40.

Delbecq, A.L., et al., *Group Techniques in Program Planning: A Guide to the Nominal Group and Delphi Processes*, Glenview : Scott, Foresman, 1975.

Shea, Michael P. et John H. Lewko, «Use of a Stakeholder Advisory Group to Facilitate the Utilization of Evaluation Results», *Canadian Journal of Program Evaluation*, Vol. 10, N° 1, avril-mai 1995, p. 159 à 162.

Uhl, Norman, et Carolyn Wentzel, «Evaluating a Three-day Exercise to Obtain Convergence of Opinion», *Canadian Journal of Program Evaluation*, Vol. 10, N° 1, avril-mai 1995, p. 151 à 158.

**Notes :**

## 4.7 Études de cas

Lorsqu'un programme se compose d'une série de projets ou de cas, l'évaluateur peut avoir recours à une série d'études de cas «particuliers» pour évaluer et expliquer ses résultats. Comme la consultation de spécialistes, les études de cas sont une forme d'enquête en soi dont l'importance justifie que nous y consacrons une section.

Les études de cas sont utilisées pour évaluer les résultats d'un programme au moyen d'un examen approfondi plutôt qu'étendu de cas ou de projets précis. Contrairement aux techniques de collecte de données décrites jusqu'ici, elles font ordinairement appel à une combinaison de diverses méthodes; on s'en sert généralement lorsqu'il est impossible, pour des raisons budgétaires ou pratiques, de constituer un échantillon suffisamment gros, ou lorsqu'il faut avoir des données très détaillées.

Normalement, les études de cas portent sur un certain nombre de cas ou de projets précis à partir desquels l'évaluateur espère tirer des renseignements portant sur l'ensemble du programme. Il est donc très important de choisir judicieusement les cas, afin que les conclusions qu'on en tire puissent s'appliquer à l'ensemble de la population cible. Malheureusement, il est fréquent que les cas soient choisis de façon peu scientifique (ou qu'on n'en choisisse pas assez), au point qu'il est impossible d'en tirer des inférences statistiques valides.

Il peut aussi arriver qu'on choisisse un cas parce qu'on le considère comme critique, voire comme le plus représentatif. Pourtant, s'il aboutit à de mauvais résultats, on risque de mettre sérieusement en doute l'efficacité de l'ensemble du programme, indépendamment des résultats des autres cas étudiés. Nous reviendrons plus loin sur ces deux situations, à savoir les cas les plus représentatifs et les cas critiques.

Supposons qu'on ne puisse déterminer les résultats d'un programme de subventions à un secteur d'activité industrielle donné qu'en se fondant sur un examen détaillé des états financiers des entreprises et sur des entrevues exhaustives auprès des gestionnaires, des comptables et des techniciens intéressés. Avec de telles exigences, il serait extrêmement coûteux d'utiliser un gros échantillon. Par conséquent, l'évaluateur pourrait opter pour un petit échantillon de cas qu'il considérerait comme représentatifs de l'ensemble de la population. Pourtant, il ne lui serait possible de généraliser ses résultats à toute la population en supposant des circonstances semblables dans des cas n'ayant pas fait l'objet d'étude. Il n'est donc pas toujours facile de poser une hypothèse comme celle-là, car elle peut être contestée ou mise en doute, ce qui risque de saper la crédibilité des conclusions.

Lorsqu'il s'agit de mesurer les résultats d'un programme, l'étude d'un cas critique peut être plus défendable que celle d'un échantillon représentatif. Prenons l'exemple d'une entreprise qui aurait reçu presque tous les fonds d'un programme conçu pour réaliser un projet industriel. L'évaluation de l'incidence de la subvention sur le projet (en a-t-elle provoqué la mise en œuvre, et, si oui, quels en ont été les

avantages) peut grandement contribuer à la mesure des résultats de l'ensemble du programme. Il s'ensuit que l'étude d'un cas critique peut être un outil aussi valable qu'important d'évaluation d'un programme.

Toutefois, dans le contexte d'une évaluation, on se sert plus souvent des études de cas pour comprendre de quelle façon le programme a été mis en œuvre et pourquoi certaines choses se sont produites que pour prendre des mesures précises.

Plus souvent qu'autrement, les résultats ne sont pas aussi évidents qu'on l'aurait prévu. L'évaluateur peut prétendre qu'ils sont attribuables à des «interactions complexes», à des «variables accessoires» ou tout simplement à une «variance inexpliquée», alors qu'on a tout simplement négligé un facteur important à l'étape de l'étude préparatoire. Cela risque de se produire assez souvent, puisqu'on connaît rarement d'avance le processus qui lie les intrants aux extrants et aux résultats. Néanmoins, il est relativement important de le connaître, et l'évaluateur peut y arriver grâce à des méthodes de collecte de données susceptibles de donner une idée de l'imprévu, dont assurément celle des études de cas.

En réalité, on peut avoir recours à des études de cas à bien des fins, y compris les suivantes :

- explorer les nombreuses conséquences d'un programme;
- sensibiliser l'évaluateur au contexte dans lequel le programme se déroule;
- faire ressortir les «variables accessoires» pertinentes;
- évaluer les conséquences à long terme du programme (Alkin, 1980).

### **Avantages et inconvénients**

Les études de cas permettent à l'évaluateur de réaliser une analyse plus approfondie que les méthodes plus générales.

C'est probablement l'atout le plus important des études de cas, étant donné que, dans la pratique, l'ampleur de l'analyse réalisable avec des méthodes globales est souvent limitée. La profondeur de l'analyse réalisable grâce aux études de cas fait que leurs résultats sont souvent très utiles. De plus, ces études peuvent mener à des hypothèses explicatives qui favorisent une analyse plus poussée.

Les études de cas sont généralement coûteuses et longues à exécuter, et c'est pourquoi on ne peut normalement pas analyser un échantillon de cas statistiquement fiable. Par conséquent, une série d'études de cas n'a habituellement pas la base statistique nécessaire pour qu'on puisse généraliser les conclusions qu'on en tire.



L'analyse approfondie que les études de cas favorisent nécessite généralement d'importantes ressources et de longs délais, ce qui limite le nombre de celles qu'on peut réaliser. On ne s'attend donc pas, normalement, à en tirer des résultats susceptibles d'être généralisés sur le plan statistique. Leur principal rôle consiste à donner un aperçu général du déroulement du programme et à favoriser sa compréhension. C'est pour cette raison qu'on recommande habituellement qu'elles soient faites avant (ou en même temps) les autres méthodes de collecte de données, dont les conclusions sont plus généralisables.

### Références : Études de cas

Campbell, D.T., «Degrees of Freedom and the Case Study», *Comparative Political Studies*, Vol. 8, 1975, p. 178 à 193.

Campbell, D.T. et J.C. Stanley, *Experimental and Quasi-experimental Designs for Research*, Chicago : Rand-McNally, 1963.

Cook, T.D. et C.S. Reichardt, *Qualitative and Quantitative Methods in Evaluation Research*, Thousand Oaks : Sage Publications, 1979, chapitre 3.

Favaro, Paul et Marie Billinger, «A Comprehensive Evaluation Model for Organizational Development», *Canadian Journal of Program Evaluation*, Vol. 8, N° 2, octobre-novembre 1993, p. 45 à 60.

Maxwell, Joseph A., *Qualitative Research Design: An Interactive Approach*, Thousand Oaks : Sage Publications, 1996.

McClintock, C.C., *et al.*, «Applying the Logic of Sample Surveys to Qualitative Case Studies: The Case Cluster Method», in Van Maanen, J., éd., *Qualitative Methodology*, Thousand Oaks : Sage Publications, 1979.

Yin, R., *The Case Study as a Rigorous Research Method*, Thousand Oaks : Sage Publications, 1986.

### Notes :

## 4.8 Résumé

Nous avons analysé dans le présent chapitre six méthodes de collecte de données utilisées pour l'évaluation de programmes : le dépouillement de la documentation spécialisée, l'étude de dossiers, les observations directes, les enquêtes, la consultation de spécialistes et les études de cas.

Les deux premières de ces méthodes consistent à recueillir des données secondaires, et les quatre autres, des données primaires. Pour faciliter l'analyse et la compréhension, nous les avons présentées séparément. Toutefois, dans le contexte de l'évaluation de programmes, il faudrait les utiliser ensemble pour appuyer les stratégies de recherche retenues.

Le dépouillement de la documentation et l'étude de dossiers sont indispensables à l'évaluation. Il faudrait y voir dès l'étape de l'étude préparatoire, ainsi que dans les premières phases de l'évaluation proprement dite. Ce sont des démarches qui servent à définir le contexte du programme à l'étude ainsi qu'à proposer des façons plausibles d'attribuer les résultats observés à un programme donné. Elles peuvent aussi permettre à l'évaluateur d'éviter de recueillir des données superflues, en lui indiquant ou en mettant en évidence des données pertinentes ou équivalentes qui existent déjà ailleurs.

Un grand nombre de méthodes étudiées dans le présent chapitre sont utilisées pour recueillir des données sur les attitudes. L'évaluateur devrait quand même se rappeler que les attitudes changent avec le temps sous l'influence de facteurs contextuels et qu'elles sont subjectives. Par exemple, une enquête consistant à interroger des gens sur les résultats d'un programme donne tout au plus à l'évaluateur l'*opinion* générale de la population visée quant aux résultats du programme. Cette opinion peut servir ou non à déterminer les résultats réels. Cela dit, la meilleure façon d'interpréter les données sur les attitudes consiste à les situer dans leur propre contexte historique et socio-économique, et c'est pourquoi il faut obtenir ces renseignements-là pour appuyer l'analyse en bonne et due forme des données sur les attitudes.

L'évaluateur devrait aussi être conscient de la subjectivité éventuelle des données obtenues grâce à certaines méthodes de collecte, surtout l'observation directe, la consultation de spécialistes, voire parfois les études de cas. Ce n'est pas nécessairement un inconvénient, mais il est nécessaire de bien établir la validité externe de toutes les conclusions. D'un autre côté, rien ne vaut ces méthodes de collecte pour obtenir des données holistiques en profondeur sur l'effet d'un programme. Combinées avec des données quantitatives, les données qualitatives sont un outil très efficace lorsqu'on veut vérifier le rapport entre un programme et ses résultats.

Enfin, comme il est rare qu'une seule méthode de collecte de données soit entièrement satisfaisante pour l'évaluation d'un programme quelconque, il est préférable d'en combiner plusieurs et de puiser ses données à différentes sources, tout en respectant les contraintes, bien entendu.

## Chapitre 5

# MÉTHODES ANALYTIQUES

### 5.1 Introduction

Les méthodes analytiques utilisées pour l'évaluation devraient être clairement exposées à l'étape du choix du modèle. Il ne faudrait jamais recueillir de données à moins que l'évaluateur sache exactement comment elles seront utilisées dans l'analyse. Un bon modèle d'évaluation tiendra compte de trois éléments : les questions à l'étude, les méthodes d'analyse et les données susceptibles d'être recueillies. Toutes ces pièces doivent se combiner parfaitement avant que l'évaluation commence.

Dans ce chapitre, nous allons décrire les méthodes analytiques utilisées dans l'administration fédérale pour déterminer les résultats d'un programme. Notre démarche est axée sur l'utilisation de ces méthodes en tant qu'éléments d'une stratégie d'évaluation donnée. Évidemment, elles peuvent aussi être utiles pour d'autres parties de l'évaluation. Par exemple, l'étude préparatoire comporte habituellement une analyse exploratoire qui contribue à cerner les questions à l'étude et à définir les méthodes de recherche les plus utiles. En outre, l'analyse permet d'intégrer les constatations obtenues grâce aux différentes stratégies d'évaluation.

Dans les pages qui suivent, nous allons décrire à la fois l'analyse de la mesure directe des répercussions des programmes ainsi que celle qui fait appel aux mesures de ces répercussions directes pour produire une estimation de diverses retombées des programmes. On distingue deux types de méthodes d'analyse directe, les méthodes statistiques et les méthodes non statistiques. Nous allons compléter ces descriptions avec celle de diverses méthodes d'analyse indirecte.

### 5.2 Analyse statistique

L'analyse statistique implique la manipulation de données (catégoriques) quantitatives ou qualitatives en vue de décrire des phénomènes et de procéder à des inférences quant aux relations entre variables. Les données en question peuvent être soit objectives et «concrètes», soit subjectives et «abstraites», mais les unes et les autres doivent être décrites ou organisées de façon systématique. Presque toutes les études analytiques font appel à l'analyse statistique, mais son emploi exige de la compétence et une compréhension des hypothèses sous-jacentes.

L'analyse statistique a deux raisons d'être, la première consistant à faire une **description**, ce pourquoi on utilise des tableaux statistiques afin de présenter des données quantitatives et qualitatives de façon aussi succincte que révélatrice. La seconde raison d'être des modèles statistiques consiste à faire des **inférences** pour vérifier les rapports entre les variables étudiées ou pour généraliser des constatations en les appliquant à une population plus étendue (d'après l'échantillon).

Pour faire rapport des constatations d'une étude d'évaluation, il faut souvent présenter succinctement une grande quantité de données. Les statistiques, présentées sous forme de tableau ou de graphique et de «statistiques» (comme la moyenne ou la variance) peuvent faire ressortir les principales caractéristiques des données.

Pour illustrer l'utilisation de l'analyse statistique descriptive, prenons l'exemple d'un programme d'enseignement de la langue seconde pour lequel on a évalué les connaissances des immigrants avant et après leur participation. Le tableau 3 contient deux exemples (A et B) de présentation sommaire des résultats aux examens des participants. Les deux sont des résumés descriptifs des données. Le second exemple (B) est plus ventilé (moins succinct) que le premier (A), dans lequel on présente la note moyenne (c.-à-d. la moyenne arithmétique des résultats). Cette statistique peut correspondre à un résultat moyen sans contenir de précisions sur l'étendue ou la distribution des résultats. Comme on peut le constater, la note moyenne des 43 personnes qui ont suivi tout le programme a été de 64,7 p. 100, comparativement à une note moyenne avant le programme de 61,2 p. 100.

**Tableau 3****Exemple de statistiques descriptives****A) Présentation des résultats moyens**

	Résultat moyen	Nombre de personnes ayant passé l'examen
Examen antérieur au programme	61,2	48
Examen postérieur au programme	64,7	43

**B) Présentation de la distribution des résultats**

	0-20	21-40	41-60	61-80	81-100	N
Examen antérieur au programme	6 (12,5 %)	5 (10,4 %)	8 (16,7 %)	24 (50 %)	5 (10,4 %)	48 (100 %)
Écart type = 22,6						
Examen postérieur au programme	5 (11,6 %)	5 (11,6 %)	6 (14,0 %)	20 (46,5 %)	7 (16,3 %)	43 (100 %)
Écart type = 23,7						

Par contre, dans le second exemple (B), la distribution générale des notes est présentée à partir des mêmes données brutes que celles utilisées pour le premier exemple (A). Ainsi, à l'évaluation antérieure au programme, six des participants avaient obtenu une note de 0 à 20 % et 20 autres une note de 61 à 80 %. La distribution des notes peut aussi être exprimée en pourcentages : on voit ainsi que 50 p. 100 (24/48) des participants évalués avant le programme avaient obtenu une note variant entre 61 et 80 %, alors que 16,3 p. 100 (7/43) de ceux qui l'ont été après le programme ont obtenu une note entre 81 et 100 %. Cette présentation en pourcentages fournit aussi des descriptions plus globales des données (par exemple, on constate que 60,4 p. 100 des participants évalués avant le programme ont obtenu plus de 60 % à l'examen.

Enfin, une statistique telle que l'écart type peut servir à résumer l'étendue de la distribution. L'écart type correspond à la mesure dans laquelle les résultats individuels se rapprochent de la moyenne arithmétique, c'est-à-dire de la normale. Plus l'écart type est petit par rapport à la normale, moins la distribution est étendue.

Les tableaux ne sont pas la seule façon de présenter des statistiques descriptives. On peut aisément présenter des données et des statistiques sous forme de graphiques. Les graphiques à barre sont utilisés pour les distributions, tandis que les graphiques circulaires ou les boîtes illustrent des proportions relatives. Ces présentations visuelles, faciles à produire avec des logiciels statistiques, peuvent être très utiles pour résumer des données statistiques, puisqu'elles sont souvent plus faciles à lire qu'un tableau et n'exigent pas nécessairement une compréhension de tous les aspects des statistiques pour en tirer une information utile.

Comme nous l'avons déjà indiqué, les données subjectives (fondées sur les attitudes) peuvent être traitées de la même façon que les données objectives. Supposons qu'on demande aux participants à un programme de formation d'évaluer leurs progrès sur une échelle de 1 à 5. Les résultats pourraient être présentés comme suit :

	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>Nombre</b>
<b>Nombre de répondants</b>	16	38	80	40	26	200
<b>Pourcentage</b>	8 %	19 %	40 %	20 %	13 %	
<b>Résultat moyen : 3,1</b>						

Dans ce cas-ci, on voit que 40 des 200 répondants (20 p. 100) ont évalué leurs progrès à 4 sur 5. La moyenne était de 3,1. Bien sûr, on peut contester la fiabilité et la validité de cette technique de mesure, mais il n'en reste pas moins que l'évaluateur peut s'en servir pour résumer succinctement les 200 réponses grâce à une simple analyse statistique descriptive.

**La deuxième principale raison de l'analyse statistique consiste à faire des inférences**, c'est-à-dire à tirer des conclusions sur des rapports entre variables, puis à généraliser ces conclusions pour les appliquer dans d'autres situations. Dans l'exemple du tableau 3, si nous supposons que les personnes qui ont subi des examens avant et après leur participation au programme sont un échantillon d'une population plus nombreuse, il faut déterminer si l'amélioration des résultats est réelle et attribuable au programme (ou à d'autres facteurs accessoires), ou si elle est simplement attribuable aux éléments aléatoires de l'échantillon, autrement dit à une erreur d'échantillonnage. Or, grâce à des méthodes statistiques comme l'analyse de la variance, il est possible de déterminer si les résultats moyens sont statistiquement différents.

À cet égard, il convient de souligner que tout ce qui est établi dans ce cas, est un rapport, à savoir que le résultat obtenu après la participation au programme est supérieur à celui qui l'avait été avant. Pour conclure que cette amélioration est attribuable au programme, il faut tenir compte des obstacles à la validité interne qui ont été analysés aux chapitres 2 et 3. Les vérifications statistiques telles que l'analyse

de la variance montrent simplement qu'il existe une différence statistiquement significative entre le résultat obtenu avant le programme et celui constaté après. Les vérifications statistiques ne prouvent donc pas que la différence est attribuable au programme. D'autres vérifications statistiques et des données supplémentaires peuvent aider à répondre aux questions d'attribution.

Prenons un autre exemple de rapports établis entre des variables grâce à une analyse statistique, soit celui des données présentées au tableau 4. Nous y voyons les résultats (en pourcentages) obtenus avant et après la participation au programme par des hommes et des femmes. Ces statistiques descriptives peuvent révéler les effets différents d'un programme pour divers groupes de participants. Ainsi, la première partie du tableau 4 montre que l'écart entre les résultats avant et après le programme est minime pour les hommes. Il s'ensuit que les descriptions laissent entendre que le programme a eu des effets différents selon le groupe de participants. Ces différences peuvent être des indices importants qu'il conviendrait de mener d'autres vérifications pour déterminer leur importance statistique.

Lorsqu'on étudie les données présentées aux tableaux 3 et 4, on voit que l'évaluateur pourrait avoir recours à l'analyse statistique par inférence pour estimer la force du rapport apparent, à savoir que les femmes ont obtenu de meilleurs résultats que les hommes. Des méthodes statistiques telles que l'analyse de régression (ou l'analyse loglinéaire) pourraient servir à établir l'importance de la corrélation entre les variables à l'étude. Dans ce cas-ci, le rapport entre les résultats, la participation ou la non-participation au programme et le sexe du participant pourrait être déterminé. En effet, les techniques statistiques de ce genre peuvent contribuer à déterminer l'importance des rapports entre les résultats d'un programme et les caractéristiques de ses participants.

Il est à noter que, même si les techniques statistiques dont nous venons de traiter (comme l'analyse de régression) sont souvent associées à l'analyse statistique par inférence, de nombreuses statistiques descriptives sont aussi produites dans ce contexte. L'évaluateur devrait établir une distinction entre le procédé arithmétique associé par exemple à l'estimation d'un coefficient de régression et la méthode à utiliser pour en évaluer l'importance. Il s'agit dans le premier cas d'une description et dans le second d'une inférence. Cette distinction est particulièrement importante lorsqu'on utilise un logiciel statistique pour produire de nombreuses statistiques descriptives. En effet, l'évaluateur doit faire des inférences appropriées à partir de ces statistiques-là.

<b>Tableau 4</b>					
<b>Autres données descriptives</b>					
<b>Distribution des résultats selon le sexe</b>					
<b>HOMMES</b>					
	0-20	21-40	41-60	61-80	81-100
Examen passé avant le programme	13 %	15 %	38 %	20 %	14 %
Examen passé après le programme	13 %	14 %	33 %	22 %	18 %
<b>FEMMES</b>					
Examen passé avant le programme	10 %	16 %	32 %	32 %	10 %
Examen passé après le programme	8 %	4 %	23 %	42 %	23 %

L'analyse statistique peut aussi servir à généraliser à une population plus nombreuse des constatations associées à un groupe donné. Il se peut par exemple que les résultats moyens obtenus aux examens avant et après la participation au programme qui sont présentés au tableau 3 soient représentatifs de l'ensemble de la population des immigrants, à condition qu'on ait utilisé des techniques d'échantillonnage appropriées ainsi que des méthodes statistiques acceptables pour établir les estimations. Si le groupe évalué était suffisamment important et statistiquement représentatif de l'ensemble de la population des immigrants, on devrait pouvoir s'attendre à obtenir des résultats semblables si le programme devait prendre de l'ampleur. Bien exécutée, l'analyse statistique peut donc grandement améliorer la validité externe des conclusions.

Les méthodes statistiques varient selon le *niveau des mesures* appliqué aux données (catégorique, ordinal, intervalle et rapport) ainsi que selon le *nombre de variables* en jeu. Les **méthodes paramétriques** sont fondées sur l'hypothèse que les données sont dérivées d'une population ayant une distribution normale (ou une autre distribution quelconque). D'autres méthodes «robustes» permettent toutefois à l'évaluateur de s'écarter fortement des hypothèses de normalité. Par exemple, on peut



utiliser un grand nombre de **méthodes non paramétriques (sans distribution)** pour les données ordinales.

Les **méthodes à variable unique** portent sur le rapport statistique entre une variable et une autre, alors que les **méthodes à plusieurs variables** sont conçues pour déterminer le rapport entre une ou plusieurs variables et un autre ensemble d'au moins deux variables.

Les méthodes à plusieurs variables peuvent être utilisées par exemple pour dégager des tendances, faire des comparaisons justes, préciser des comparaisons et étudier l'influence marginale d'une variable (les effets des autres facteurs restant constants).

On distingue deux types de méthodes à plusieurs variables, celles qui sont fondées sur le modèle linéaire général (c.-à-d. le modèle paramétrique normal), et celles, plus modernes, qu'on utilise pour l'analyse de plusieurs variables de données catégoriques, comme l'analyse loglinéaire. On peut aussi les répartir en deux catégories :

- a) méthodes servant à l'**analyse de la dépendance**, telles que les méthodes de régression (y compris l'analyse de la variance/covariance), la représentation fonctionnelle, l'analyse de trajectoires, les méthodes à série temporelle et à contingences multiples, ainsi que les méthodes qualitatives (catégoriques) et mixtes;
- b) méthodes utilisées pour l'**analyse de l'interdépendance**, telle que l'analyse typologique, l'analyse des composantes principales, la corrélation canonique et les analogues catégoriques.

## Avantages et inconvénients

L'analyse statistique permet de résumer les constatations d'une évaluation de façon claire, précise et fiable; c'est aussi une technique valide de détermination de la valeur statistique attribuable aux conclusions que l'évaluateur tire des données.

En dépit de ses nombreux avantages, l'analyse statistique présente plusieurs inconvénients.

Il faut être compétent pour réaliser une bonne analyse statistique.

L'évaluateur devrait consulter un statisticien professionnel aux étapes de la conception et de l'analyse de son évaluation. Il faudrait éviter de se laisser séduire par la facilité apparente de la manipulation statistique à l'aide de logiciels standard.

Les résultats d'un programme ne peuvent pas tous être analysés par des méthodes statistiques.

Par exemple, les réponses à une question ouverte sur les résultats d'un programme peuvent renfermer de longues descriptions des avantages et des répercussions néfastes du programme, alors qu'il peut être très difficile de classer — et plus encore de quantifier — ces réponses d'une façon qui se prête à l'analyse statistique sans perdre des nuances importantes, quoique subtiles.

La façon de classer les données peut tout aussi bien fausser les résultats que révéler d'importantes différences.

Même lorsque l'évaluateur a obtenu des données quantitatives, il devrait interpréter avec soin les résultats des analyses statistiques. Ainsi, les données figurant au tableau 3 pourraient être représentées d'une façon différente, comme on le voit au tableau 5. Au départ, les données sont identiques, mais les résultats présentés au tableau 5 semblent révéler un effet beaucoup plus marqué que ceux du tableau 3. Cet exemple montre bien l'importance d'utiliser des méthodes statistiques supplémentaires pour vérifier la solidité des rapports apparents. En d'autres termes, avant de conclure que les différences apparentes entre le tableau 3 ou le tableau 5 sont des résultats du programme, il faudrait pousser plus loin l'analyse statistique inférentielle.

**Tableau 5**

**Exemple de statistiques descriptives**

**A) Présentation des résultats moyens**

Examen antérieur au programme	58,4
Examen postérieur au programme	69,3

**B) Présentation de la distribution des résultats**

	0-35	36-70	71-100	N
Examen antérieur au programme	10	28	10	48 (100 %)
Examen postérieur au programme	6	11	26	43 (100 %)

Les praticiens qui se servent de l'analyse statistique doivent connaître aussi bien les hypothèses sur lesquelles la technique statistique employée est fondée que ses limites.

Une des grandes difficultés d'utilisation des méthodes analytiques, c'est que leur validité est fonction des hypothèses fondamentales qu'elles posent sur les données. Compte tenu de la grande disponibilité de logiciels statistiques, on court toujours le risque que les techniques utilisées fassent appel à des données qui doivent présenter certaines caractéristiques que les données auxquelles on a accès n'ont pas.

Bien entendu, cela peut mener à des conclusions injustifiées. Par conséquent, il est essentiel que l'évaluateur connaisse les limites des techniques qu'il emploie.

Les méthodes statistiques à plusieurs variables sont particulièrement vulnérables à ce genre d'abus, même si elles peuvent donner l'impression d'avoir bien été utilisées. Pour que lesdites méthodes soient acceptables, il faut que le modèle causal sous-jacent soit correctement spécifié.

La régression à plusieurs variables peut notamment faire tomber l'évaluateur dans les pièges suivants :

- fournir tant d'explications qu'une différence réelle est éliminée;
- ajouter des éléments superflus à un schéma simple;
- susciter un optimisme exagéré quant à la solidité des rapports de causalité établis à partir des données;
- utiliser une méthode analytique incorrecte.

### Références : Analyse statistique

Behn, R.D. et J.W. Vaupel, *Quick Analysis for Busy Division Makers*, New York : Basic Books, 1982.

Casley, D.J. et K. Kumar, *The Collection, Analysis and Use of Monitoring and Evaluation Data*, Washington (DC) : Banque mondiale, 1989.

Fienberg, S., *The Analysis of Cross-classified Categorical Data* (2<sup>e</sup> édition), Cambridge (MA) : Massachusetts Institute of Technology (MIT), 1980.

Hanley, J.A., «Appropriate Uses of Multivariate Analysis,» *Annual Review of Public Health*, Palo Alto (CA) : Annual Reviews Inc., 1983, p. 155 à 180.

Hanushek, E.A. et J.E. Jackson, *Statistical Methods for Social Scientists*, New York : Academic Press, 1977.

Hoaglin, D.C., *et al.*, *Data for Decisions*, Cambridge (MA) : Abt Books, 1982.

Morris, C.N. et J.E. Rolph, *Introduction to Data Analysis and Statistical Inference*, Englewood Cliffs (NJ) : Prentice Hall, 1981.

Ragsdale, C.T., *Spreadsheet Modelling and Decision Analysis*, Cambridge, (MA) : Course Technology Inc., 1995.

**Notes :**

### 5.3 Analyse de l'information qualitative

L'analyse non statistique est surtout appliquée à des données qualitatives, telles que les descriptions détaillées des dossiers administratifs ou des journaux d'observation sur le terrain, les affirmations directes en réponse à des questions ouvertes, la transcription de discussions en groupe et les observations de toutes sortes dont il a été brièvement question aux sections 4.1 et 4.4 à 4.7. Dans la présente section, nous nous bornerons à une description succincte de l'analyse non statistique. Pour obtenir des précisions à ce sujet, le lecteur est prié de consulter les références citées à la fin de la section.

L'analyse de données qualitatives, qui se fait ordinairement de pair avec l'analyse statistique et d'autres types d'analyses de données quantitatives, peut donner un aperçu holistique des phénomènes étudiés dans le contexte de l'évaluation. La collecte et l'analyse de l'information qualitative sont souvent «naturalistes» et fondées sur des déductions. Au début de l'étape de collecte des données ou de l'analyse, l'évaluateur ne s'appuie sur aucune théorie particulière à l'égard des phénomènes à l'étude. (Un autre type d'analyse non statistique de données quantitatives est décrit à la section 5.5 qui porte sur l'utilisation de modèles.)

Il est possible que l'analyse non statistique de données fasse davantage appel au jugement professionnel de l'évaluateur que d'autres méthodes, comme l'analyse statistique. Il s'ensuit qu'en plus de devoir bien connaître les questions qui font l'objet de l'évaluation, l'évaluateur qui effectue une analyse non statistique doit être conscient des nombreux biais qui sont susceptibles de fausser ses constatations.

Il y a plusieurs types d'analyse non statistique, dont l'analyse du contenu, l'analyse des études de cas, l'analyse inductive (y compris l'établissement de typologies) et l'analyse logique. Toutes ces méthodes sont censées faire ressortir des constantes, des thèmes, des tendances et des «motifs» des données, en plus de fournir des interprétations et des explications de ces constantes et de ces autres éléments. L'analyse des données devrait évaluer la fiabilité et la validité des constatations, par exemple grâce à une étude des hypothèses contradictoires, et elle devrait aussi analyser les cas «déviants» ou exceptionnels et faire une «triangulation» en comparant des données tirées de plusieurs sources ou obtenues grâce à d'autres méthodes de collecte et d'analyse.

Les quatre principales décisions à prendre dans le contexte d'une analyse non statistique de données portent sur la méthode analytique (résumé qualitatif, comparaison qualitative ou analyse statistique descriptive ou à plusieurs variables), sur le niveau de l'analyse, sur le moment auquel il convient de la faire, ce qui suppose des décisions quant à l'enregistrement et au codage des données ainsi qu'à l'opportunité de les quantifier, et enfin sur la façon d'intégrer l'analyse non statistique à l'analyse statistique connexe.

Bien que l'analyse non statistique (et statistique) des données suit normalement leur collecte, les deux peuvent se faire simultanément. Cette façon de procéder peut permettre à l'évaluateur de poser de nouvelles hypothèses qu'il peut vérifier aux étapes ultérieures de la collecte des données, ainsi que de cerner et de corriger d'éventuelles difficultés à cet égard, de même que d'obtenir l'information qui semble faire défaut dans les données recueillies au début. Par contre, les conclusions fondées sur une analyse hâtive risquent de biaiser la collecte ultérieure des données, voire de provoquer un changement prématuré de la conception ou de l'exécution du programme, ce qui rend bien délicate l'interprétation des constatations fondées sur toute la gamme des données recueillies.

Il est préférable de combiner l'analyse non statistique des données avec une analyse statistique de données connexes (quantitatives ou qualitatives). À cette fin, il faudrait concevoir l'évaluation de façon à ce que les deux sortes d'analyses pour lesquelles on utilise des données différentes mais connexes s'appuient, ou du moins s'éclairent mutuellement.

### **Avantages et inconvénients**

Les principaux avantages de l'analyse non statistique des données consistent à rendre possible l'examen de nombreuses questions et notions difficiles à quantifier, favorisant une approche plus holistique.

En outre, l'analyse non statistique permet à l'évaluateur de tirer profit de toute l'information disponible. Il se peut que les constatations tirées d'une analyse de ce genre soient plus détaillées que celles tirées d'une analyse purement statistique.

Toutefois, les conclusions fondées uniquement sur une analyse non statistique risquent de ne pas être aussi crédibles que d'autres fondées sur de l'information et des données provenant de sources multiples, ainsi que sur plusieurs méthodes d'analyse.

La validité et l'exactitude des conclusions d'une analyse non statistique sont fonction de la compétence et du jugement de l'évaluateur; leur crédibilité dépend de la logique des arguments présentés.

Cook et Reichardt (1979), Kidder et Fine (1987) ainsi que Pearsol (1987), entre autres, ont étudié ces questions de façon plus détaillée.

**Références : Analyse non statistique de l'information qualitative**

- Cook, T.D. et C.S. Reichardt, *Qualitative and Quantitative Methods Evaluation Research*, Thousand Oaks : Sage Publications, 1979.
- Guba, E.G., «Naturalistic Evaluation,» in Cordray, D.S., *et al.*, éd., *Evaluation Practice in Review*, Vol. 34 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1987.
- Guba, E.G. et Y.S. Lincoln, *Effective Evaluation : Improving the Usefulness of Evaluation Results Through Responsive and Naturalistic Approaches*, San Francisco : Jossey-Bass, 1981.
- Krueger, R.A., *Focus Groups : A Practical Guide for Applied Research*, Thousand Oaks : Sage Publications, 1988.
- Levine, M., «Investigative Reporting as a Research Method : An Analysis of Bernstein and Woodward's *All the President's Men*», *American Psychologist*, Vol. 35, 1980, p. 626 à 638.
- Miles, M.B. et A.M. Huberman, *Qualitative Data Analysis : A Sourcebook of New Methods*, Thousand Oaks : Sage Publications, 1984.
- Nachmias, C. et D. Nachmias, *Research Methods in the Social Sciences*, New York : St. Martin's Press, 1981, chapitre 7.
- Patton, M.Q., *Qualitative Evaluation Methods*, Thousand Oaks : Sage Publications, 1980.
- Pearsol, J.A., éd., «Justifying Conclusions in Naturalistic Evaluations», *Evaluation and Program Planning*, Vol. 10, N° 4, 1987, p. 307 à 358.
- Rossi, P.H. et H.E. Freeman, *Evaluation : A Systematic Approach* (2<sup>e</sup> édition), Thousand Oaks : Sage Publications, 1989.
- Van Maasen, J., éd., *Qualitative Methodology*, Thousand Oaks : Sage Publications, 1983.
- Webb, E.J., *et al.*, *Nonreactive Measures in the Social Sciences* (2<sup>e</sup> édition), Boston : Houghton Mifflin, 1981.
- Williams, D.D., éd., *Naturalistic Evaluation*, Vol. 30 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1987.

**Notes :**

## 5.4 Analyse des autres résultats des programmes

Les évaluations ont généralement pour objet de mesurer les résultats directs des programmes. Or, il arrive fréquemment que les programmes aient des répercussions plus générales ou à plus long terme qui présentent elles aussi de l'intérêt. On analyse fréquemment ces répercussions en transformant les résultats directs mesurés de façon à les déterminer. Au chapitre 1, nous avons distingué trois types de résultats d'un programme, à savoir :

- les extrants (qui sont de nature opérationnelle);
- les retombées intermédiaires (y compris les avantages pour les clients du programme et, parfois, les inconvénients imprévus pour le client et pour d'autres personnes);
- les retombées définitives (qui sont étroitement liées aux objectifs du programme et habituellement aux objectifs globaux du gouvernement, c'est-à-dire les avantages économiques, l'amélioration de la santé, de la sécurité et du mieux-être).

Dans les analyses de ce genre, on utilise normalement un modèle analytique conçu pour transposer les résultats des deux premiers types en résultats du troisième (ou en résultats différents du deuxième) :

Activités du programme	⇒	Extrants opérationnels/ Avantages pour les clients	⇒	Avantages pour les clients/ retombées
------------------------	---	---	---	--

Prenons un cas bien simple d'application de cette méthode, celui du programme d'enseignement de la lecture aux immigrants qui est censé améliorer leurs perspectives d'emploi. L'enchaînement logique du programme est présenté graphiquement de la façon suivante :

Programme d'enseignement de la lecture	⇒	Accroissement des compétences en lecture	⇒	Augmentation des revenus/ meilleures perspectives d'emploi
--	---	--	---	---

Dans un cas comme celui-là, on emploierait une stratégie d'évaluation visant à déterminer l'effet incrémentiel du programme d'enseignement de la lecture sur les compétences à cet égard, puis on prendrait des mesures. On utiliserait ensuite un modèle préétabli pour transformer les changements observés des compétences en lecture des participants en résultats escomptés pour leurs revenus et leurs perspectives d'emploi : les améliorations observées en ce qui concerne les compétences en lecture seraient donc transformées en retombées pour les perspectives d'emploi et les revenus,



*le tout étant fondé sur des recherches antérieures qui ont un établi un lien entre ces variables et les compétences en lecture.*

Il faut observer que toutes les analyses de ce genre sont des solutions de rechange à l'évaluation directe des résultats généraux d'un programme. Dans notre exemple, l'évaluateur pourrait mesurer directement les retombées du programme en ce qui concerne la capacité des participants d'obtenir des emplois mieux rémunérés. Il pourrait notamment se servir d'un modèle quasi expérimental pour comparer un groupe de participants au programme avec un groupe témoin afin de déterminer si les premiers ont augmenté leurs revenus d'emploi comparativement aux membres du second. Cela dit, les méthodes plus indirectes peuvent toutefois se révéler préférables pour de nombreuses raisons.

L'analyse de résultats généraux rend possible l'estimation des retombées à long terme.

Les effets secondaires ne sont pas souvent immédiats, et les contraintes de l'évaluation ne permettraient pas toujours d'assurer un suivi sur une longue période.

L'analyse des résultats généraux permet à l'évaluateur de déterminer des retombées qui sont difficiles à mesurer directement.

Il peut être extrêmement difficile ou complexe d'évaluer directement les résultats généraux, particulièrement dans le cadre d'un projet d'évaluation donné. D'une certaine façon, ces méthodes réduisent les risques qui se posent pendant l'évaluation. En effet, lorsqu'on mesure d'abord les résultats immédiats, on peut avoir confiance qu'au moins certains d'entre eux auront été mesurés d'une manière valide. Par contre, en allant directement aux résultats généraux, qui peuvent se révéler difficiles à mesurer, on risque de se retrouver sans aucune mesure valide.

L'analyse des résultats généraux est utile pour l'évaluation de retombées générales déjà étudiées.

En raison des difficultés de mesure que nous venons de décrire, l'évaluateur pourrait être tenté d'utiliser un rapport entre les effets à court terme et les résultats généraux d'un programme qui ont été déterminés grâce à des recherches antérieures (si, bien sûr, on dispose d'une telle recherche). Par exemple, dans le cas du programme d'enseignement de la lecture, il est vraisemblable qu'on ait déjà fait des recherches poussées afin d'explorer le rapport entre les compétences en lecture, les perspectives d'emploi et les revenus. En pareil cas, l'évaluateur pourrait se fonder sur les résultats de ces recherches, en décidant d'axer sa stratégie d'évaluation sur la mesure des améliorations des compétences en lecture résultant du programme; l'augmentation des revenus des participants qui s'ensuivrait vraisemblablement serait alors une conclusion qui a déjà été prouvée par des recherches antérieures.

## Notes :

## 5.5 Utilisation de modèles

Toutes les évaluations servant à établir que certains résultats découlent des activités d'un programme sont fondées sur un modèle implicite ou explicite. Sans théorie sur la façon dont le programme produit des résultats observés, l'évaluateur travaillerait à l'aveuglette et serait incapable de lui attribuer des résultats de façon crédible. Cela ne signifie toutefois pas que les modèles doivent être complètement structurés dès le début du travail d'évaluation. Généralement, ils sont révisés et améliorés en cours de route, à mesure que l'équipe d'évaluation développe ses connaissances.

Les diverses disciplines des sciences sociales ont tendance à adopter des approches quelque peu différentes face aux modèles, mais il reste quand même de nombreux points communs.

Les modèles que nous allons décrire dans cette section sont les suivants :

- modèles de simulation;
- modèles d'entrées-sorties;
- modèles micro-économiques;
- modèles macro-économiques;
- modèles statistiques.

### 5.5.1 Modèles de simulation

La simulation peut s'avérer utile pour les évaluateurs. Toute transformation des intrants du programme en extrants peut être exposée sur une feuille de calcul et modélisée par un évaluateur ayant une certaine formation et un peu de pratique.

L'évaluateur a souvent recours à un modèle quantitatif explicite parce que ses données sont incertaines. Lorsqu'il doit traiter des intervalles de variation plutôt que des chiffres, en jonglant avec les probabilités, il peut lui être extrêmement utile de pouvoir simuler les probabilités d'extrants ou de résultats. Dans les années 1990, le progiciel qui a amélioré les capacités de simulation des tableurs électroniques a offert cette capacité à de nombreux évaluateurs qui auraient peut-être eu moins tendance à opter pour des approches quantitatives, dans d'autres conditions.

Un modèle de simulation peut transformer des intrants en résultats. Prenons par exemple un programme des Douanes aux postes frontaliers qui sont aménagés au bord des autoroutes, et disons qu'on a formulé une nouvelle série de questions à poser aux points d'entrée. L'administration du nouveau questionnaire prend en moyenne 11 secondes de plus que celle de l'ancien. On pourrait utiliser un modèle pour évaluer ses conséquences sur le temps d'attente des clients.

Un modèle de simulation comporte essentiellement trois composantes, soit des intrants, un modèle mathématique et des extrants. On utilise surtout deux types de modèles mathématiques, les *stochastiques*, qui font intervenir des variables aléatoires, et les *déterministes*, qui ne contiennent pas de variables de ce genre.

À certains égards, les modèles de simulation ressemblent aux autres méthodes statistiques, comme à l'analyse de régression, qui sont d'ailleurs susceptibles d'être utilisées pour les établir. Une fois établi, le modèle de simulation traite les intrants comme des données qu'il doit utiliser plutôt que des éléments sur lesquels il doit se fonder. Le modèle mathématique génère des extrants qui peuvent être comparés aux résultats réels.

Les évaluateurs s'intéressent de plus en plus à un modèle de simulation donné, à savoir le modèle de risque fondé sur une feuille de calcul coûts-avantages. Lorsque les intrants du modèle coûts-avantages sont représentés par des approximations et des probabilités (plutôt que comme des données certaines), le modèle de risque produit des données sur les prochaines valeurs et sur les probabilités du résultat essentiel (habituellement la valeur actualisée nette). Ces données peuvent être très utiles pour un gestionnaire qui tente d'évaluer le degré de risque d'un programme, ou pour un évaluateur appelé à faire une évaluation du seuil de tolérance et du risque (voir la section 5.6, Analyse coûts-avantages et analyse coût-efficacité).

### **Avantages et inconvénients**

Le principal avantage des modèles de simulation est qu'il permet à l'évaluateur d'estimer les effets incrémentiels dans des situations complexes et incertaines. Par contre, leur principal inconvénient est d'exiger une excellente compréhension de la dynamique du programme ainsi qu'une certaine maîtrise de l'établissement de modèles quantitatifs.

Il faudrait également noter que les modèles de simulation peuvent fournir de l'information valable *ex ante*, soit de l'information sur les répercussions éventuelles d'un mode d'action donné avant sa réalisation. De l'information de ce type peut assurément être fort utile avant d'exclure des solutions de rechange indésirables. *Ex post*, les répercussions réelles d'un nouveau programme ou des changements apportés à un programme existant sont mieux évaluées par les méthodes empiriques, comme une analyse de régression ou les modèles présentés au chapitre 3.

### **Références : Modèles de simulation**

Buffa, E.S. et J.S. Dyer, *Management Science Operations Research : Model Formulation and Solution Methods*, New York : John Wiley and Sons, 1977.

Clemen, R.T., *Making Hard Decisions*. Duxbury Press, 1991, sections 1 à 3.

Ragsdale, C.T., *Spreadsheet Modelling and Decision Analysis*, Cambridge (MA) : Course Technology Inc., 1995.

### **Notes :**

### 5.5.2 Modèles d'entrées-sorties

Un modèle d'entrées-sorties est un modèle économique statique conçu pour décrire l'interdépendance mutuelle de différentes parties d'une économie. Dans ce contexte, l'économie est considérée comme un système d'activités interdépendantes, c'est-à-dire agissant directement et indirectement les unes sur les autres. Le modèle d'entrées-sorties est utilisé pour décrire la façon dont un secteur utilise comme intrants des extrants d'autres secteurs, et vice versa. C'est donc une déconstruction systématique de l'économie qui décrit l'échange de biens et de services nécessaires à la fabrication de produits finis (biens et services).

Ce genre de modèle peut être utilisé pour dériver des prévisions multisectorielles qui sont intrinsèquement cohérentes avec les tendances économiques, ainsi que des évaluations quantitatives détaillées des effets secondaires directs et indirects d'un programme quelconque, ou de toute combinaison de programmes. Plus précisément, le modèle d'entrées-sorties peut produire une description détaillée de l'effet d'un programme gouvernemental sur la production et la consommation actuelles de biens et de services.

La structure des entrées de chaque secteur de production est expliquée en fonction de sa technologie. Le modèle précise les «coefficients techniques» correspondant à la quantité de biens et de services, y compris la main-d'œuvre, dont le secteur a besoin pour produire une unité d'extrait. Il précise aussi un ensemble de «coefficients de capital» correspondant à l'ensemble des bâtiments, du matériel et des stocks nécessaires à la transformation de la combinaison voulue d'intrants en extrants. Les caractéristiques de la consommation définissent la demande d'intrants (le revenu, par exemple) de tous les secteurs de production de l'économie, y compris les ménages. On peut donc analyser ces caractéristiques, de même que la production et la consommation de n'importe quel bien ou service.

Pour démontrer l'utilité d'un modèle d'entrées-sorties, il suffit d'imaginer l'effet de mesures fiscales sélectives (hypothétiques) sur l'emploi dans le secteur des télécommunications. Supposons que ces mesures fiscales assurent un traitement préférentiel au secteur et influent donc directement sur la quantité, la composition et le prix de ses extrants, lesquels influent à leur tour sur sa demande et sur son utilisation de main-d'œuvre. Le modèle fait appel à des coefficients correspondant à l'état actuel de la technologie de pointe et à des équations permettant de préciser la consommation et la production attendues de chaque secteur.

Au départ, on commence par estimer l'importance des changements résultant de l'application des mesures fiscales sélectives, en se fondant sur les valeurs de la consommation et de la production prévues du matériel de télécommunication. Le modèle d'entrées-sorties peut ensuite utiliser comme intrant l'augmentation de la consommation de ce matériel, en produisant comme extrait l'accroissement estimatif de la main-d'œuvre du secteur des télécommunications résultant des mesures fiscales.

## Avantages et inconvénients

Autrefois, on utilisait plus fréquemment les modèles d'entrées-sorties dans les économies à planification centrale. Ces modèles, ponctuels et statiques, sont essentiellement descriptifs et, par conséquent, ils ne sont pas très efficaces pour inférer des effets probables liés aux politiques pour l'avenir.

Malheureusement, on a fréquemment mal utilisé les modèles de ce genre dans les évaluations. Le pire exemple est celui de l'analyse des dépenses de programme dans un secteur afin d'estimer les «effets» supposés qui en auraient résulté, sans tenir compte de l'atténuation des effets négatifs qui sont causés par les mesures fiscales ou les emprunts contractés pour financer le programme.

En outre, dans une économie en pleine évolution, ces modèles présentent un autre inconvénient majeur, puisqu'ils ne tiennent pas nécessairement compte des changements des coefficients de production attribuables au progrès technologique, ni des changements relatifs des prix des intrants. Par conséquent, lorsque ces changements se produisent, le modèle d'entrées-sorties décrit une composition incorrecte des intrants d'un secteur donné, ce qui entraîne des estimations incorrectes des résultats supplémentaires du programme étudié. À cet égard, soulignons que le modèle d'entrées-sorties de Statistique Canada est inévitablement fondé sur des données datant d'un certain nombre d'années, et que, en tant que macro-modèle, il n'est pas particulièrement bien adapté à la description des effets des petites dépenses typiques de la plupart des programmes.

## Références : Modèles d'entrées-sorties

Canada, Statistique Canada, *La structure par entrées-sorties de l'économie canadienne 1961-1981*, Ottawa, avril 1989, n° de cat. 15-201F.

Chenery, H. et P. Clark, *Inter-industry Economics*, New York : John Wiley and Sons, 1959.

Leontief, W., *Input-output Economics*, New York : Oxford University Press, 1966.

## Notes :

### 5.5.3 Modèles micro-économiques

Les modèles micro-économiques décrivent le comportement économique d'unités économiques individuelles (personnes, ménages, entreprises ou autres organisations) fonctionnant dans une structure de marché et dans des circonstances données. Comme la plupart des programmes sont dirigés exactement à ce niveau, ces modèles peuvent être extrêmement utiles pour l'évaluateur. Ils sont fondés sur le système des prix et normalement représentés par des équations correspondant aux fonctions de l'offre et de la demande d'un bien ou d'un service. Ces équations décrivent le rapport entre le prix et l'extrait, et il est souvent possible d'en faire une représentation graphique avec des courbes de l'offre et de la demande.

Le rendement des modèles micro-économiques est limité par un certain nombre d'hypothèses. Par exemple, on suppose toujours que les consommateurs se comportent de façon à maximiser leur degré de satisfaction et ce, d'une façon rationnelle. Les spécialistes se servent des modèles micro-économiques pour modéliser le comportement du marché, les combinaisons optimales des intrants, le comportement des consommateurs en fonction des coûts et les niveaux de production optimaux.

Dans la pratique, on peut avoir recours à des modèles micro-économiques pour estimer les résultats d'un programme dans la mesure où les prix et les extraits peuvent en décrire les effets. La figure 4 est un exemple d'un modèle micro-économique permettant de décrire l'effet qu'un programme de taxe d'accise sur les cigarettes aurait sur le revenu des fabricants ou sur le tabagisme chez les adolescents.

D'après la figure 4, le prix et la quantité de cigarettes produites et consommées avant l'imposition de la taxe d'accise correspondraient respectivement à  $P_0$  et  $Q_0$ . La taxe d'accise ferait augmenter le coût des cigarettes et cette augmentation serait représentée dans le modèle micro-économique par une courbe de l'offre croissante. Le nouveau prix serait donc plus élevé et la nouvelle production plus faible qu'avant l'imposition de la taxe d'accise. À ce moment-là, les recettes de l'industrie des cigarettes équivalaient à  $P_0 \times Q_0$ , mais depuis, avec la nouvelle taxe d'accise, elles sont tombées à  $P_1 \times Q_1$ . Cette baisse des recettes des fabricants de cigarettes par suite de l'imposition de la taxe d'accise serait fonction de la pente des courbes de l'offre et de la demande qui est elle-même déterminée par plusieurs facteurs.

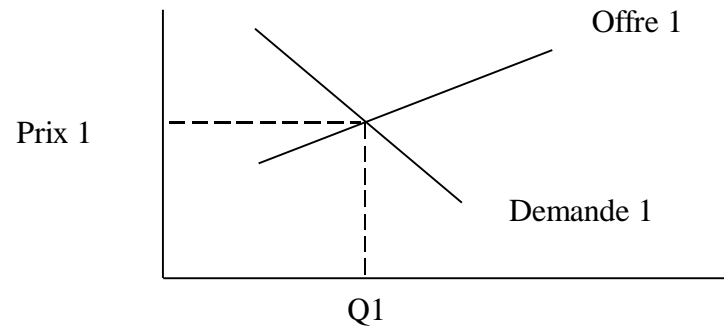
#### Avantages et inconvénients

Il faut normalement avoir recours à un économiste pour établir un modèle micro-économique des effets d'un programme, mais cela en vaut souvent la peine, puisque ces modèles peuvent apporter beaucoup d'information sur la raison d'être d'un programme et fournir une base pour mesurer ses effets et son efficacité.

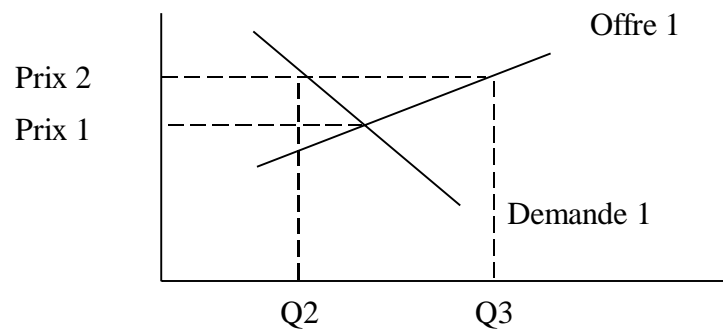
Figure 4

## Modèle de l'effet d'une taxe d'accise

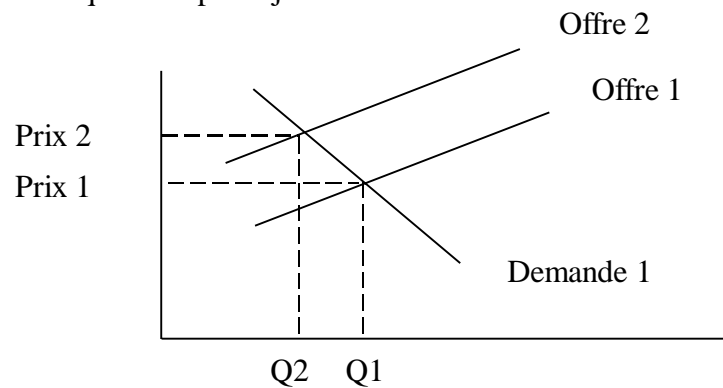
A) Avant l'imposition de la taxe



B) Effet initial de la taxe (offre supérieure à la demande)



C) Nouvel équilibre après ajustement des fabricants



**Références : Modèles micro-économiques**

Henderson, J. et R. Quandt, *Micro-economic Theory*, New York : McGraw-Hill, 1961.

Polkinghorn, R.S., *Micro-theory and Economic Choices*, Richard Irwin Inc., 1979.

Samuelson, P., *Foundations of Economic Analysis*, Cambridge (MA) : Harvard University Press, 1947.

Watson, D.S., *Price Theory in Action*, Boston : Houghton Mifflin, 1970.

**Notes :**



### 5.5.4 Modèles macro-économiques

Les modèles macro-économiques sont essentiellement utilisés pour des études sur l'inflation, le chômage et les sujets faisant appel à d'importants ensembles de données, comme le produit national brut. On s'en sert pour tenter d'expliquer et de prédire les rapports entre ces variables.

Ce sont des modèles utiles parce qu'ils révèlent les retombées économiques — une amélioration de la production, du revenu ou de l'emploi ou encore une hausse des taux d'intérêt ou de l'inflation — les plus susceptibles de découler de l'application d'une politique ou de l'exécution d'un programme monétaire et financier.

Voici un exemple d'utilisation d'un modèle macro-économique : supposons qu'un évaluateur cherche à évaluer les retombées sur l'emploi d'un programme gouvernemental de subvention de certains types d'exportation et que les effets du programme sur les ventes à l'exportation ont déjà été mesurés. Les données sur l'accroissement incrémentiel de ces ventes seraient introduites dans un modèle macro-économique de l'économie canadienne qui pourrait alors estimer les retombées du programme sur l'emploi.

#### Avantages et inconvénients

Le modèle macro-économique a l'avantage de préciser les liens critiques entre les variables générales globales. En outre, il permet de broser un tableau général qu'on peut ensuite utiliser pour comparer des programmes canadiens à des programmes analogues mis en œuvre dans d'autres pays (à condition que les hypothèses et les critères de validité du modèle demeurent intacts).

Pour l'évaluation des résultats d'un programme, le modèle macro-économique présente toutefois de graves inconvénients. En effet, il peut aboutir à des résultats erronés si l'on omet des facteurs clés. En outre, ses données des intrants sont habituellement dérivées d'un autre modèle plutôt que directement mesurées, ce qui ajoute un autre élément d'incertitude à l'analyse.

Enfin, dans bien des cas, la valeur prédictive, surtout à court terme, du modèle macro-économique laisse vraiment à désirer. Néanmoins, c'est un outil qu'on peut utiliser avec profit si les retombées dérivées à l'étude sont à long terme et si l'évaluation porte sur un programme important pour l'économie.

#### Références : Modèles macro-économiques

Gordon, R.A., *Economic Instability and Growth : The American Record*, Harper & Row, 1974.

Heilbroner, R.L. et L.C. Thurow, *Economics Explained*, Toronto : Simon and Schuster Inc., 1987.

Nelson, R., Merton, P. et E. Kalachek, *Technology, Economic Growth and Public Policy*, Washington (DC) : Brookings Institute, 1967.

Okun, A., *The Political Economy of Prosperity*, Norton, 1970.

Silk, L., *The Economists*, New York : Avon Books, 1976.

**Notes :**

### 5.5.5 Modèles statistiques

Les études d'évaluation font appel à beaucoup de types de modèles statistiques dont le plus simple est une présentation de données relatives à une seule variable organisée de façon à en illustrer la configuration. Les tableaux de corrélation de deux variables sont l'instrument de base de l'analyse et du rapport d'évaluation. En fait, même les données analysées à l'aide d'autres modèles sont souvent présentées dans des tableaux de corrélation, pour les rendre plus transparentes et plus accessibles aux décideurs que celles des modèles plus complexes.

Habituellement, les programmes cliniques (dans les domaines de la santé et de l'éducation, par exemple) sont basés sur de petits échantillons, de sorte que l'évaluateur doit utiliser des modèles «d'analyse de la variance» pour en préciser les effets. À l'inverse, les programmes destinés à une grande partie de la population (subventions au commerce ou programmes d'emploi, par exemple) génèrent normalement de vastes ensembles de données et on peut donc avoir recours alors à des «modèles linéaires» d'analyse de régression pour en déterminer les effets. La plupart des programmes du gouvernement fédéral sont de ce dernier type, et c'est pourquoi nous allons nous concentrer sur eux dans cette section.

L'analyse de régression peut servir à vérifier une relation hypothétique, à établir des relations entre des variables qui sont susceptibles d'expliquer les résultats d'un programme, à cerner les cas inhabituels (valeurs aberrantes) qui dévient des normes ou à faire des prévisions sur les retombées futures d'un programme. Il s'agit là d'une technique parfois exploratoire (pour concocter des rapports approximatifs), mais on l'emploie plus souvent comme confirmation et mesure finale d'une relation causale entre le programme et ses effets constatés. De fait, il est important que le modèle de régression se fonde sur un raisonnement *a priori* au sujet de la causalité. Il faudrait éviter de rechercher des données au hasard, au risque d'obtenir des résultats sans valeur, et c'est pourquoi il faut s'efforcer de spécifier et de calibrer le modèle en utilisant seulement la moitié des données disponibles pour ensuite déterminer sa capacité de prédiction des résultats révélés par l'autre moitié des données. S'il est un bon prédicteur, le modèle est probablement robuste.

Il faut se rappeler que la corrélation n'implique pas nécessairement un rapport de causalité. Par exemple, deux variables peuvent être simplement corrélées simplement parce qu'elles sont toutes deux causées par une troisième variable. Ainsi, on peut établir une corrélation entre la température diurne élevée et le nombre de prêts agricoles consentis parce que les deux se produisent surtout en été, mais cela ne veut pas dire que les prêts agricoles sont consentis parce qu'il fait chaud durant la journée.

L'analyse de régression tend aussi à inverser le rapport de causalité; c'est d'ailleurs une de ses difficultés reconnues. On peut observer, par exemple, que les entreprises qui obtiennent des stimulations d'incitation d'un programme d'aide au commerce extérieur augmentent leurs ventes à l'exportation. Or, cela peut s'expliquer simplement du fait que les entreprises qui ont de grosses ventes à l'étranger sont plus crédibles que les autres, et qu'il leur est donc plus facile d'obtenir des subventions.

On pourrait aussi dire que ce sont leurs ventes à l'étranger qui font obtenir des subventions aux entreprises, plutôt que l'inverse.

Les modèles statistiques ont souvent une importance cruciale pour la détermination des effets incrémentiels. Par exemple, Santé Canada pourrait utiliser un modèle épidémiologique pour préciser les effets de sa Stratégie nationale sur le sida, tandis que le ministère des Finances Canada pourrait utiliser un modèle des revenus pour estimer les effets fiscaux d'un régime éventuel d'aide à la famille. Pour arriver à constituer de tels modèles, il faut généralement une connaissance approfondie du secteur de programmes analysé, ainsi qu'une maîtrise de la technique statistique utilisée.

### **Avantages et inconvénients**

Les modèles statistiques sont polyvalents. Bien construits, ils fournissent des estimations très utiles des résultats d'un programme. Toutefois, ils doivent être bien spécifiés et validés si l'on veut que les résultats soient fiables, ce qui n'est pas toujours aussi facile qu'on pourrait le croire à prime abord.

En outre, l'évaluateur n'arrive pas toujours à faire des inférences à partir d'un modèle statistique. Il se peut par exemple que le modèle porte uniquement sur certains groupes d'âge, ou seulement sur des personnes de certaines régions, auquel cas il est souvent impossible, à partir des résultats, d'en généraliser les effets éventuels à d'autres groupes d'âge ou à d'autres régions.

### **Références : Modèles statistiques**

Chatterjee, S. et B. Price, *Regression Analysis by Example* (2<sup>e</sup> édition), New York : John Wiley and Sons, 1995.

Fox, J., *Linear Statistical Models and Related Methods, with Applications to Social Research*, New York : John Wiley and Sons, 1984.

Huff, D., *How to Lie with Statistics*, Penguin, 1973.

Jolliffe, R.F., *Common Sense Statistics for Economists and Others*, Routledge and Kegan Paul, 1974.

Mueller, J.H., *Statistical Reasoning in Sociology*, Boston : Houghton Mifflin, 1977.

Sprent, P., *Statistics in Action*, Penguin, 1977.

### **Notes :**

## 5.6 Analyse coûts-avantages et analyse coût-efficacité

Tous les programmes visent à générer des avantages qui l'emportent sur leurs coûts. Après avoir estimé les divers coûts et avantages résultant du programme, l'évaluateur peut comparer les deux pour déterminer si le programme est valable. Les deux méthodes les plus fréquemment utilisées à cette fin sont l'analyse coûts-avantages et l'analyse coût-efficacité. Généralement, on s'en sert pour obtenir des renseignements sur la valeur actualisée nette d'un programme. Dans l'analyse coûts-avantages, les avantages du programme sont exprimés en termes monétaires et comparés à ses coûts, alors que, dans l'analyse coût-efficacité, les résultats du programme, exprimés en unités non monétaires — par exemple le nombre de vies sauvées — sont comparés à ses coûts exprimés en dollars.

À l'étape de la planification, on peut mener des analyses coûts-avantages et coût-efficacité *ex ante* (avant coup) en se fondant sur des estimations des coûts et des avantages escomptés. La plupart des ouvrages et des publications sur l'analyse coûts-avantages la considèrent comme un instrument d'analyse *a priori*, et surtout comme un moyen d'examiner les avantages nets d'un projet ou d'un programme proposé nécessitant des investissements ou des immobilisations considérables (voir par exemple Mishan, 1972; Harberger, 1973; Layard, 1972; Sassone et Schaffer, 1978 et Schmid, 1989).

Lorsqu'un programme fonctionne depuis un certain temps, on peut aussi avoir recours à une analyse coûts-avantages ou coût-efficacité *ex post* (après coup) pour déterminer si les coûts *réels* du programme sont justifiés par ses avantages *réels*. Pour une étude plus détaillée de l'utilisation de l'analyse coûts-avantages dans le contexte de l'évaluation, voir Thompson (1980) ou Rossi et Freeman (1989). Il y a aussi un aperçu de cette méthode dans le *Guide de l'analyse avantages-coûts* (1997), une publication du Conseil du Trésor, ainsi que dans les études de cas connexes.

L'analyse coûts-avantages consiste à comparer les avantages tangibles et intangibles d'un programme à ses coûts directs et indirects. Après avoir cerné et mesuré (ou estimé) les avantages et les coûts, on les transforme pour les exprimer en termes communs, habituellement monétaires, de façon à pouvoir les comparer en calculant la valeur actualisée nette du programme. Quand les coûts et les avantages sont étalés dans le temps, il faut les actualiser pour les ramener à une année commune avec le taux d'actualisation approprié.

Pour faire une analyse de ce genre, il faut d'abord choisir le point de vue à partir duquel les coûts et les avantages du programme seront calculés. **On en reconnaît habituellement trois, soit le point de vue de la personne, le point de vue financier du gouvernement fédéral et le point de vue social (pour l'ensemble du Canada).** Les coûts et les avantages d'un programme varient généralement selon le point de vue. Le plus courant pour les analyses avantages-coûts dans l'administration fédérale est le **point de vue social**, qui tient compte de tous les coûts et avantages pour la société. Toutefois, le point de vue de la personne et le point de vue financier du gouvernement peuvent contribuer à faire ressortir des perspectives différentes sur la valeur du programme ou encore à expliquer les raisons de sa réussite ou de son échec.

Rossi et Freeman (1989) ont produit une analyse plus approfondie des différences entre les trois points de vue.

On part du **point de vue de la personne** pour examiner les coûts et les avantages du programme pour le participant (qui pourrait être une personne, une famille, une entreprise ou une organisation sans but lucratif). Les analyses coûts-avantages pour lesquelles on adopte ce point de vue aboutissent souvent à des rapports avantages-coûts élevés, parce que le gouvernement ou la société subventionnent le programme dont le participant bénéficie.

D'un autre côté, lorsque l'analyse est effectuée du **point de vue financier du gouvernement fédéral**, les coûts et les avantages sont évalués du point de vue de la source du financement. Il s'agit essentiellement d'une analyse financière dans laquelle on examine les coûts financiers et les avantages financiers directs pour l'État. Les flux de trésorerie qu'on étudierait normalement dans ce contexte comprendraient les coûts d'administration du programme, les sorties de fonds directes (les subventions), les taxes et impôts perçus par le gouvernement (notamment l'impôt sur le revenu des sociétés, l'impôt sur le revenu des particuliers, les taxes de vente fédérale et autres droits), la réduction des prestations d'assurance-chômage ou d'assurance-emploi et les changements éventuels des paiements de péréquation et de transfert.

Par contre, pour l'**analyse coûts-avantages du point de vue social**, on part du point de vue de l'ensemble de la société, de sorte que l'analyse est à la fois plus exhaustive et plus difficile, puisqu'il faut tenir compte des *résultats* généraux du programme, et que les prix du marché, qui sont un bon indicateur des coûts et des avantages pour la personne ou pour une organisation (l'État) risquent de ne pas refléter fidèlement la valeur réelle de ces deux variables pour la société. Ils peuvent être faussés, par exemple, en raison des subventions ou des taxes et impôts. Même s'ils ressemblent à ceux qui sont utilisés dans les analyses du point de vue du particulier et de celui du gouvernement, les éléments examinés dans l'analyse coûts-avantages du point de vue social sont appréciés et calculés différemment (voir Weisbrod *et al.*, 1980). Par exemple, les *coûts d'opportunité* pour la société sont différents de ceux qu'assume un participant au programme. En outre, les paiements de transfert sont exclus des coûts dans le contexte d'une analyse coûts-avantages du point de vue social, puisqu'ils doivent aussi être considérés comme des avantages pour la société et que les deux s'annulent par conséquent.

Les analyses coûts-avantages faites du point de vue du gouvernement ou du point de vue social tendent à produire des rapports avantages-coûts inférieurs à ceux des analyses analogues qui sont réalisées du point de vue de la personne parce que l'État ou la société assument généralement la totalité du coût du programme, alors que la personne, elle, peut bénéficier de tous ses avantages, en n'assumant qu'une fraction infime du coût total. Néanmoins, les analyses coûts-avantages des programmes gouvernementaux devraient être faites du point de vue social.

Pour sa part, **l'analyse coût-efficacité** exige aussi que les coûts et les avantages du programme étudié soient quantifiés, quoique les avantages (ou les effets) ne sont pas alors exprimés en dollars. Il s'agit plutôt de combiner les données sur les effets ou l'efficacité du programme aux données sur ses coûts de façon à pouvoir comparer le coût et l'efficacité du programme. Par exemple, dans une analyse coût-efficacité, on exprimerait les résultats d'un programme d'éducation en parlant de la progression moyenne d'un niveau de lecture (données sur les résultats) par tranche de 1 000 \$ (données sur les coûts) investis dans le programme. Les avantages (effets) sont exprimés en termes quantitatifs — mais pas en dollars — dans l'analyse coût-efficacité.

Ce genre d'analyse est fondé sur les mêmes principes que l'analyse coûts-avantages. Les hypothèses utilisées, par exemple pour le calcul des coûts et l'actualisation, sont les mêmes dans les deux cas. Au fond, l'analyse coût-efficacité permet de comparer et de classer des programmes en fonction du coût pour atteindre certains buts. Les données sur l'efficacité peuvent être combinées avec celles sur les coûts pour déterminer l'efficacité maximale correspondant à un coût donné, ou encore le coût le plus bas permettant d'atteindre un degré d'efficacité particulier.

Les données qui sont nécessaires à l'exécution d'analyses coûts-avantages et coût-efficacité peuvent provenir de diverses sources. Bien entendu, les recherches dans les dossiers détaillés des programmes devraient générer beaucoup d'informations sur les coûts, et ces données peuvent souvent être complétées grâce à des sondages auprès des bénéficiaires. D'autre part, les données sur les avantages peuvent être recueillies par n'importe quelle des autres méthodes dont nous avons déjà parlé dans cette publication.

Supposons par exemple qu'on a entrepris une évaluation pour vérifier l'hypothèse qu'un programme de santé mentale rejetant l'hospitalisation en lui préférant la prestation de soins de santé dans la collectivité serait plus efficace que la méthode de traitement prévalant à l'heure actuelle, et supposons aussi qu'on a employé un modèle expérimental pour obtenir une estimation des effets incrémentiels de ce programme innovateur. Dès que les effets incrémentiels seraient connus, l'analyse coûts-avantages pourrait permettre de les évaluer et de les comparer aux coûts.

### **Avantages et inconvénients**

La documentation sur les avantages et les inconvénients de l'analyse coûts-avantages et de l'analyse coût-efficacité abonde (voir par exemple Greer et Greer, 1982, ainsi que Nobel, 1977). Nous nous contenterons ici de faire valoir succinctement un certain nombre de points à cet égard.

L'analyse coûts-avantages porte sur la valeur nette d'un programme.

Il ne s'agit pas, en l'occurrence, d'estimer des avantages et des coûts précis d'un programme, mais plutôt de les résumer de façon qu'on puisse juger et comparer des solutions de rechange. Il faut mesurer dans un autre contexte le degré auquel les objectifs ont été atteints, en faisant appel à un autre modèle d'évaluation et à des méthodes différentes de collecte des données. Par la suite, les données sur les résultats du programme peuvent être utilisées comme intrants pour les analyses globales coûts-avantages et coût-efficacité.

L'évaluateur doit aborder la question de l'attribution ou des effets incrémentiels avant de réaliser une analyse coûts-avantages.

Par exemple, de 1994 à 1997, le gouvernement fédéral a mis en œuvre un programme d'infrastructures à frais partagés avec les municipalités et les provinces. Avant de pouvoir analyser les coûts et les avantages de ce programme ou de ses solutions de rechange, il faudrait établir des mesures des effets incrémentiels afin de déterminer jusqu'à quel point le programme a changé ou accéléré les travaux d'infrastructure municipaux. C'est seulement après avoir déterminé les effets incrémentiels qu'on peut raisonnablement passer à l'évaluation et à la comparaison des coûts et des avantages.

Les analyses coûts-avantages et coût-efficacité aident souvent l'évaluateur à déterminer tous les coûts et tous les résultats d'un programme.

À elles seules, les analyses coûts-avantages et coût-efficacité ne suffisent pas à expliquer des effets et des résultats particuliers.

Ces techniques ne permettent pas de déterminer pourquoi un objectif donné n'a pas été atteint, ni pourquoi un effet particulier s'est produit. Toutefois, comme elles comparent systématiquement les avantages et les coûts, elles sont utiles puisqu'elles fournissent des renseignements valides aux décideurs.

Ces analyses comportent de nombreuses difficultés méthodologiques.

Il est souvent difficile d'exprimer en dollars les avantages et les coûts d'un programme. Il peut être très malaisé d'attribuer une valeur monétaire à des résultats dans les domaines de l'éducation, de la santé (quelle valeur attribuer à la vie humaine ou encore à sa qualité), voire de l'équité et de la répartition du revenu. Toutes les évaluations de cet ordre sont et resteront toujours très discutables. En outre, même lorsqu'on réussit à les exprimer en dollars, les coûts et les avantages doivent être actualisés à un point commun dans le temps afin qu'on puisse les comparer. Les auteurs traitant des analyses coûts-avantages sont loin de l'unanimité à ce sujet. Ils continuent à discuter du taux d'actualisation optimal. Dans son *Guide de l'analyse avantages-coûts*, le Conseil du Trésor recommande à l'évaluateur de faire une analyse de risque (simulation), avec une fourchette de taux se situant autour de 10 p. 100 par année, compte tenu de l'inflation.



L'évaluateur devrait toujours faire une analyse de sensibilité des hypothèses sous-jacentes aux analyses coûts-avantages et coût-efficacité, pour vérifier la solidité des résultats obtenus.

Compte tenu des hypothèses qu'il faut poser pour comparer les avantages et les coûts d'un programme, l'évaluateur aurait intérêt à effectuer une analyse de sensibilité afin de déterminer dans quelle mesure ses conclusions sont fonction de chacune de ses hypothèses. En outre, il devrait s'efforcer de vérifier à quel degré ces conclusions varient lorsque les hypothèses changent. Si les résultats de l'analyse dépendent largement de la valeur d'un intrant donné, il peut valoir la peine de supporter le coût d'études supplémentaires pour vérifier cette valeur. Soulignons que, contrairement à certains autres types de méthodes d'évaluation, l'analyse coût-efficacité permet à l'évaluateur d'effectuer une analyse de sensibilité à la fois systématique et rigoureuse.

On a parfois recours à l'analyse coût-efficacité lorsqu'il est trop difficile de convertir en termes monétaires les valeurs qu'elle utilise.

L'analyse coût-efficacité permet parfois à l'évaluateur de comparer et de classer les solutions de rechange mais, comme les avantages ne sont pas convertis en dollars, il est impossible de déterminer la valeur nette du programme ou de comparer des programmes différents en se fondant sur les mêmes critères.

Par contre, l'analyse coûts-avantages permet d'utiliser des techniques grâce auxquelles il est possible de comparer et d'évaluer même des coûts et des avantages qui sont difficiles à mesurer en termes monétaires. Malheureusement, elle exige souvent des ajustements délicats des mesures des coûts et des avantages en raison de l'utilisation d'hypothèses incertaines, ce qui risque d'inquiéter les gestionnaires qui craignent souvent, parfois à raison, que ces hypothèses et ces ajustements risquent de favoriser la manipulation des résultats en privilégiant n'importe quel biais éventuel de l'analyse.

De plus, la détermination des coûts et des avantages est souvent d'autant plus difficile que les ministères et organismes publics ne conservent pas à cet égard des dossiers grâce auxquels il serait facile de les comparer. Pour la plupart des programmes, les données sur les coûts que les services intéressés conservent ont trait à de nombreuses activités et sont organisées pour faciliter la tâche des administrateurs, et non celle de l'évaluateur.

### **Références : Analyse coûts-avantages et analyse coût-efficacité**

Angelsen, Arild et Ussif Rashid Sumaila, *Hard Methods for Soft Policies : Environmental and Social Cost-benefit Analysis*, Bergen, Norvège : Institut Michelsen, 1995.

Australie, ministère des Finances, *Handbook of Cost-benefit Analysis*, Canberra, 1991.

Banque mondiale, Institut de développement économique, *The Economics of Project Analysis : A Practitioner's Guide*, Washington (DC), 1991.

Belli, P., *Guide to Economic Appraisal of Development Projects*, Washington (DC) : Banque mondiale, 1996.

Bentkover, J.D., Covdlo, V.T. et J. Mumpower, *Benefits Assessment : The State of the Art.*, Dordrecht, Pays-Bas : D. Reidel Publishing Co., 1986.

Canada, Bureau du Vérificateur général, «Le choix et l'application des techniques de collecte des éléments probants en vérification d'optimisation des ressources», *Analyse coûts-avantages*, Ottawa, 1994, annexe B5.

Canada, Secrétariat du Conseil du Trésor, *Guide de l'analyse avantages-coûts*, Ottawa, 1997 (doit paraître pendant l'été 1997).

Harberger, A.C., *Project Evaluation : Collected Papers*, Chicago : Markham Publishing Co., 1973.

Miller, J.C. III et B. Yandle, *Benefit-cost Analyses of Social Regulation*, Washington : American Enterprise Institute, 1979.

Sang, H.K., *Project Evaluation*, New York : Wilson Press, 1988.

Sassone, P.G. et W.A. Schaffer, *Cost-benefit Analysis : A Handbook*, New York : Academic Press, 1978.

Schmid, A.A., *Benefit-cost Analysis : A Political Economy Approach*, Boulder : Westview Press, 1989.

Self, P., *Econocrats and the Policy Process : The Politics and Philosophy of Cost-benefit Analysis*, Londres : Macmillan, 1975.

Skaburskis, Andrejs et Fredrick C. Collignon, «Cost-effectiveness Analysis of Vocational Rehabilitation Services», *Canadian Journal of Program Evaluation*, Vol. 6, N° 2, octobre-novembre 1991, p. 1 à 24.

Skelton, Ian., «Sensitivity Analysis in Multi-criteria Decision Aids : A Demonstration of Child Care Need Assessment», *Canadian Journal of Program Evaluation*, Vol. 8, No. 1, avril-mai 1993, p. 103 à 116.

Sugden, R. et A. Williams, *The Principles of Practical Cost-benefit Analysis*, Oxford : Oxford University Press, 1978.

Thompson, M., *Benefit-cost Analysis for Program Evaluation*, Thousand Oaks : Sage Publications, 1980.

Van Pelt, M. et R. Timmer, *Cost-benefit Analysis for Non-Economists*, Institut d'économie des Pays-Bas, 1992.

Watson, Kenneth, «The Social Discount Rate», *Canadian Journal of Program Evaluation*, Vol. 7, N° 1, avril-mai 1992, p. 99 à 118.

Yates, Brian T., *Analyzing Costs, Procedures, Processes, and Outcomes in Human Services*, Thousand Oaks : Sage Publications, 1996.

**Notes :**

## 5.7 Résumé

Dans ce chapitre, nous avons décrit plusieurs méthodes d'analyse des données qui devraient faire partie intégrante de la stratégie d'évaluation, dans la pratique. Les éléments de cette stratégie devraient d'ailleurs former un tout cohérent dans lequel les questions à évaluer, le modèle, les méthodes de collecte des données et la technique d'analyse des données optimale devraient s'agencer aussi harmonieusement que possible.

Nous avons étudié une vaste gamme de méthodes d'analyse dans ce manuel, en décrivant plusieurs types d'analyses statistiques et non statistiques d'évaluation des résultats d'un programme et de méthodes d'estimation de leurs retombées, notamment grâce à l'utilisation de modèles, ainsi que des méthodes de détermination des coûts. Il sera bien sûr toujours difficile de décider quand et comment utiliser une méthode donnée, puis de le faire habilement et judicieusement.

## Chapitre 6

### CONCLUSIONS

Nous avons analysé les principaux facteurs dont l'évaluateur devrait tenir compte lorsqu'il conçoit des stratégies d'évaluation des résultats d'un programme, en concentrant notre analyse sur l'interaction entre les facteurs suivants :

- le contexte du programme et de la prise de décisions;
- la stratégie d'évaluation elle-même (modèles, collecte des données et analyse des données).

Nous avons consacré trois chapitres aux principaux aspects de l'élaboration des stratégies d'évaluation : les modèles (chapitre 3), la collecte des données (chapitre 4) et les méthodes analytiques (chapitre 5).

En l'occurrence, l'objectif est le suivant : les évaluations doivent produire des constatations et des conclusions sur le rendement du programme à la fois opportunes, pertinentes, crédibles et objectives, fondées sur une collecte et une analyse des données valides et fiables. En outre, les rapports d'évaluation devraient présenter les constatations et les conclusions clairement, de façon équilibrée, tout en précisant leur fiabilité.

C'est notamment sur ces normes que les ministères et les organismes fédéraux se fondent pour mener leurs activités internes d'auto-évaluation et d'amélioration de la qualité. À mesure que l'expérience du Canada en matière d'évaluation s'élargira et s'approfondira, il viendra, sans aucun doute, s'y ajouter d'autres normes de qualité présentant un intérêt particulier pour divers groupes d'évaluateurs canadiens et pour leurs clients.

## Annexe 1

# ENQUÊTES

Dans la section 4.5, nous avons étudié les enquêtes en tant que méthodes de collecte des données pour l'évaluation des programmes; nous avons joint à cette section une liste de références à consulter pour obtenir des renseignements supplémentaires à ce sujet. C'était nécessaire parce que la conception d'une enquête devrait normalement mettre à profit la compétence des spécialistes du domaine. Compte tenu de la fréquence d'utilisation des enquêtes dans les évaluations, nous avons jugé opportun d'ajouter la présente annexe au manuel pour y présenter une analyse plus détaillée des principaux facteurs dont il faut tenir compte dans la conception d'une enquête. **Néanmoins, cette annexe ne doit pas être considérée comme substituable à la consultation de spécialistes.**

La conception d'une enquête comporte trois volets fondamentaux : la conception de l'*échantillonnage*, le choix de la *méthode d'enquête* et l'établissement de l'*instrument de mesure*. Nous avons analysé brièvement chacun de ces volets en précisant les principaux écueils qui y sont associés.

### 1.1 Échantillonnage

Lorsqu'il n'est ni possible ni efficient d'étudier toute la population visée par le programme, il faut utiliser une méthode d'échantillonnage. La portée et la nature de cette méthode devraient satisfaire aux trois exigences suivantes.

#### **Les constatations doivent être généralisables à l'ensemble de la population visée.**

Si l'évaluateur doit présenter des conclusions au sujet de l'ensemble de la population visée en se fondant sur une enquête auprès d'un échantillon, il doit s'assurer que les constatations tirées de l'enquête seront généralisables à toute cette population. Si tel est le cas, il doit habituellement avoir recours à un échantillon aléatoire (plutôt qu'à un échantillon non aléatoire). L'évaluateur doit être très conscient du risque de biais statistiques qui se produisent normalement lorsqu'un échantillon non aléatoire est considéré comme un échantillon aléatoire et qu'on en tire des inférences injustifiées. Ces biais sont souvent attribuables à une utilisation inappropriée ou négligente des méthodes d'échantillonnage aléatoire.

### **La méthode choisie doit satisfaire aux exigences minimales de précision.**

Le degré de précision et le niveau de confiance attendus de l'enquête doivent être précisés. La théorie statistique peut fournir des estimations de l'erreur d'échantillonnage pour des échantillons de différentes tailles, autrement dit de la précision des estimations. Il s'ensuit que la taille de l'échantillon serait fonction du degré de précision recherché. L'évaluateur devrait accorder plus d'importance à la précision qu'à la taille de l'échantillon, prise isolément. Rappelons ici qu'il existe différentes formules de calcul de la taille de l'échantillon ainsi que différents types de mesures (ou d'estimations), notamment l'importance d'une caractéristique de la population et la proportion de la population dans une catégorie donnée. Il n'est pas rare qu'on utilise la mauvaise formule pour calculer la taille minimale de l'échantillon nécessaire.

### **Les coûts d'échantillonnage doivent respecter les limites du budget.**

Certaines méthodes d'échantillonnage, comme l'échantillonnage stratifié et l'échantillonnage répété, ont été conçues afin de réduire à la fois la taille de l'échantillon et le coût de la prise des mesures. À cet égard, il convient de souligner que le raffinement des méthodes d'échantillonnage peut se révéler rentable.

Après avoir posé ces trois exigences, on peut passer à l'établissement du processus d'échantillonnage, qui comprend les six étapes suivantes.

- (i) **Définir la population.** Cette définition doit être précise et détaillée; elle comprend souvent la date, le lieu et les caractéristiques socio-économiques pertinentes. Par exemple : toutes des femmes, de 18 ans et plus, habitant l'Ontario, ayant participé au programme entre le 15 et le 30 novembre 1982 et ayant actuellement un emploi.
- (ii) **Préciser la base de sondage.** La base de sondage est une liste des éléments de la population (p. ex., noms dans un annuaire téléphonique, liste d'électeurs, liste de prestataires au dossier). Si elle n'existe pas, il peut falloir la créer (totalement ou partiellement) en appliquant une stratégie d'échantillonnage.
- (iii) **Préciser l'unité d'échantillonnage.** On entend par là l'unité employée pour l'échantillonnage, comme le lieu, le pâté de maisons, le ménage ou l'entreprise.
- (iv) **Préciser la méthode d'échantillonnage.** C'est la méthode utilisée pour choisir les unités d'échantillonnage (p. ex., échantillonnage systématique ou stratifié).

- (v) **Déterminer la taille de l'échantillon.** Il s'agit alors de déterminer le nombre d'unités d'échantillonnage ainsi que le pourcentage de la population à inclure dans l'échantillon.
- (vi) **Choisir l'échantillon.**

Des erreurs attribuables à d'autres facteurs qu'à l'échantillonnage peuvent se glisser à chaque étape de ce processus. Par exemple, la population définie peut ne pas correspondre à la population cible ou la base de sondage peut ne pas coïncider exactement avec la population. En pareil cas, les mesures ou inférences affectées peuvent être biaisées, et donc trompeuses. Supposons par exemple qu'on effectue un sondage auprès des bénéficiaires de subventions dans le cadre d'une évaluation d'un programme d'aide à un secteur d'activité industrielle et que la base de sondage des entreprises se limite à celles qui ont reçu plus qu'un certain montant. Dans ces conditions, il est bien évident que toute généralisation des résultats portant sur l'ensemble des bénéficiaires de subventions ne serait pas valide si elle était fondée sur un échantillon choisi à partir de cette base.

Ces erreurs attribuables à d'autres facteurs peuvent aussi s'introduire dans presque toutes les activités d'enquête. Par exemple, les répondants peuvent interpréter différemment les questions, les proposés au traitement des résultats peuvent faire des erreurs, et il est toujours possible qu'il y ait des erreurs dans la base de sondage même. Bref, il peut y avoir des erreurs autres que celles qui sont attribuables à l'échantillonnage, aussi bien dans les enquêtes sur des échantillons que dans les recensements, alors que les erreurs d'échantillonnage ne sont possibles, bien entendu, que dans le premier de ces deux types d'enquête.

## 1.2 Méthodes d'enquête

C'est habituellement la technique de collecte des données utilisée qui détermine le genre d'enquête. Le choix de cette technique est donc extrêmement important pour toutes les enquêtes fondées sur des réponses individuelles. Nous allons maintenant analyser les trois méthodes d'enquête de base.

### Entrevues téléphoniques

Pour établir son échantillon, l'enquêteur part d'une base de sondage contenant des numéros de téléphone, choisit une unité d'échantillonnage dans cette base et réalise une entrevue téléphonique avec une personne bien précise qui répond à l'appel ou encore avec la première personne qui y répond. Il existe aussi une autre technique, dite de composition aléatoire, où l'enquêteur compose un numéro choisi au hasard sans même savoir s'il existe ou si l'abonné est une entreprise, un hôpital ou un ménage. Dans la pratique, les deux techniques sont utilisées ensemble. Par exemple, il est courant d'avoir recours à celle de la composition aléatoire pour produire une première liste de numéros. Ensuite, on choisit au hasard des numéros dans cette liste pour produire le jeu de numéros de l'échantillon.



## **Entrevues directes**

Il existe essentiellement trois façons de recueillir des renseignements grâce à des entrevues; elles se prêtent toutes bien aux entrevues directes. Même si elles sont toutes utilisables aussi pour réaliser des entrevues téléphoniques, il est extrêmement rare que l'une ou l'autre des deux premières donnent de bons résultats dans ce contexte. Chacune suppose une préparation, une conceptualisation et des instruments différents, et chacune présente des avantages et des inconvénients. Voici les trois façons de réaliser des entrevues :

- entrevues non structurées;
- entrevues fondées sur un guide;
- entrevues suivant une présentation type.

### **Entrevues non structurées**

Cette façon de procéder est entièrement fondée sur des questions posées de façon spontanée au cours de l'entrevue, souvent dans le cadre d'une observation continue des activités du programme. Dans ce genre d'entrevue, il arrive que l'interlocuteur ne se rende même pas compte qu'il est interrogé. L'avantage de cette façon de procéder est de permettre à l'évaluateur de tenir compte des différences individuelles et situationnelles; il peut personnaliser ses questions de façon à avoir un échange en profondeur avec son interlocuteur, dans une atmosphère détendue. C'est une technique particulièrement utile lorsque l'évaluateur peut la mettre à profit pour explorer le programme sur une période assez longue, ce qui lui permet de préparer ses entrevues en se fondant sur les réponses qu'il a obtenues auparavant.

Malheureusement, les entrevues non structurées ont l'inconvénient de s'étendre sur une longue période, puisqu'il faut parfois plusieurs conversations avant d'obtenir réponse à une série uniforme de questions. En outre, c'est une façon de procéder plus vulnérable que les autres aux effets et aux biais intervenant pendant l'entrevue, puisqu'elle dépend largement de l'habileté de l'enquêteur.

### **Entrevues fondées sur un guide**

Les guides d'entrevue sont des listes de questions ou de thèmes à soulever pendant l'entrevue. Ils sont conçus pour faire en sorte que les mêmes questions de base soient traitées dans toutes les entrevues et proposent à l'enquêteur des aspects ou des sujets qu'il est libre d'explorer afin d'approfondir une question donnée. Autrement dit, ce sont des cadres, dans lesquels l'enquêteur conçoit et organise ses questions et décide des points à approfondir.

Cette façon de procéder a l'avantage de permettre à l'enquêteur d'exploiter au maximum le temps limité dont il dispose. L'entrevue est plus systématique et plus complète parce que les questions à discuter sont précisées à l'avance. C'est une méthode particulièrement utile pour les entrevues de groupe, car elle permet à

l'enquêteur de faire en sorte que les participants ne s'écartent pas du sujet, tout en tenant compte des points de vue individuels.

Pourtant, c'est une façon de procéder qui présente plusieurs inconvénients. En effet, même avec un guide d'entrevue, l'enquêteur peut parfois oublier des questions importantes. La souplesse dont il dispose pour l'enchaînement et la formulation des questions peut en outre réduire nettement la comparabilité des réponses. De plus, la technique peut aussi sembler très intimidante pour l'interlocuteur, et l'impression que celui-ci se fait de l'enquêteur peut aussi saper la validité et la fiabilité des réponses.

### **Entrevues suivant une présentation type**

Lorsqu'il faut obtenir de chaque personne interrogée des renseignements strictement comparables, on peut avoir recours à une présentation type permettant à l'enquêteur de poser les mêmes questions à chacune. Avant le début des entrevues, on rédige le texte des questions ouvertes et fermées telles qu'elles seront posées. Toutes les explications et les précisions nécessaires sont formulées à l'avance dans le texte, comme d'ailleurs toutes les questions éventuelles d'exploration.

Cette méthode réduit le risque de biais de l'enquêteur, puisque celui-ci doit poser les mêmes questions à chaque répondant. L'entrevue est systématique et ne fait à peu près pas appel au jugement de l'enquêteur. En outre, l'analyse des données est facilitée, puisqu'on peut regrouper les questions et les réponses qui se ressemblent. De plus, le texte même du questionnaire peut être soumis aux décideurs avant le début des entrevues. Enfin, comme l'enquêteur doit fonctionner dans un cadre précis, les entrevues sont habituellement plus courtes avec cette méthode qu'avec les autres.

Par contre, ce genre d'entrevue ne permet pas à l'enquêteur d'approfondir les thèmes qui pourraient être soulevés seulement au cours de la conversation, même si l'emploi de questions ouvertes permet de mitiger un peu cet inconvénient. De plus, c'est une façon de procéder qui empêche jusqu'à un certain point le chercheur de tenir compte des différences individuelles et des circonstances.

### **Méthodes combinées**

Dans les études d'évaluation, la meilleure façon de procéder est souvent une combinaison de la méthode du guide d'entrevue et de l'entrevue avec présentation type. Il s'ensuit que, dans la plupart des cas, un certain nombre de questions sont formulées d'avance, quoique l'enquêteur dispose de la latitude voulue pour poser d'autres questions et pour décider quand il vaut la peine d'approfondir certains points. On utilise souvent une présentation type au début de chaque entrevue, après quoi l'enquêteur est plus libre de s'intéresser à d'autres sujets généraux pour le reste de l'entrevue.

## Enquête postale

La troisième méthode d'enquête de base consiste à envoyer le questionnaire par la poste au répondant, en l'invitant à y répondre et à le retourner à l'expéditeur. Pour obtenir les taux de réponse élevés indispensables à une bonne analyse, on utilise essentiellement des questions fermées dans la plupart des enquêtes de ce genre. C'est une méthode qui a l'avantage d'atteindre un gros échantillon de répondants à un coût relativement modique. En outre, avec des questions quantitatives fermées, l'analyse des données est relativement simple, puisqu'on peut comparer directement les réponses et les résumer et les regrouper facilement. Par contre, cette méthode présente l'inconvénient que les répondants doivent adapter leur vécu et leurs opinions pour les faire correspondre à des catégories préétablies, ce qui peut fausser ce qu'ils voulaient dire en limitant leurs choix. Pour pallier ces difficultés, on ajoute souvent des questions ouvertes afin que les répondants puissent préciser et développer leurs réponses.

Cela dit, l'un des principaux écueils associés aux enquêtes postales est leur faible taux de réponse, qui peut aussi poser un problème dans le cas des enquêtes téléphoniques et des entrevues directes, quoique dans une moindre mesure. Un faible taux de réponse peut être imputable à de nombreux facteurs, dont la non-disponibilité des répondants ou le refus de participer. On a fréquemment recours aux trois stratégies suivantes pour accroître le taux de réponse :

- relancer les non-répondants par téléphone;
- interroger les non-répondants;
- faire un suivi par la poste.

Dans le premier cas, l'enquêteur téléphone aux non-répondants — après un certain temps — pour les presser de remplir le questionnaire.

La deuxième stratégie consiste à prendre un échantillon de non-répondants pour remplir le questionnaire avec eux au cours d'une entrevue téléphonique ou directe. Ensuite, on pondère les résultats de ces entrevues afin qu'ils soient représentatifs de l'ensemble de la population des non-répondants, puis, en combinant les résultats avec ceux des répondants, on arrive à faire des généralisations non biaisées à l'ensemble de la population. Toutefois, pour que cette technique soit valide, il faut avoir établi scientifiquement l'échantillon des non-répondants sollicités.

La troisième stratégie, le suivi postal, ressemble à celle de la relance téléphonique, mais elle est habituellement moins efficace. Elle consiste à envoyer de nouveau le questionnaire aux non-répondants après un certain temps, en leur demandant de bien vouloir le remplir.

De toute évidence, il peut arriver qu'on ne puisse pas faire grand-chose pour améliorer le taux de réponse, faute de temps et d'argent. Il faut donc tenir compte du taux de non-réponse quand on tire des conclusions sur la population étudiée à partir de l'information recueillie auprès des membres de l'échantillon.

Un taux de réponse faible déforme l'estimation des résultats, étant donné qu'il est possible que les attitudes ou les intérêts des non-répondants ne correspondent pas à ceux des répondants. Heureusement, il existe plusieurs méthodes qui permettent de corriger le biais attribuable à un mauvais taux de réponse, comme le sous-échantillonnage des non-répondants.

### **Enquête sur des objets (inventaire)**

Les méthodes d'enquête que nous venons de décrire s'appliquent à des personnes, mais on peut aussi réaliser des enquêtes sur des objets, comme des immeubles, des maisons ou toutes sortes d'articles. Les principes d'échantillonnage utilisés dans le cas des personnes valent également pour les objets. En fait, l'élément le plus important d'une enquête est un enquêteur compétent, car c'est à lui de veiller à ce que des mesures appropriées soient prises, de les colliger et de les transmettre fidèlement. Dans les enquêtes sur des objets, le risque de distorsion des mesures est au moins aussi élevé que le risque de déformation attribuable aux biais de l'enquêteur dans les enquêtes basées sur des entrevues.

Prenons par exemple le cas d'un programme d'aide à un secteur d'activité industrielle conçu pour inciter les entreprises à mettre au point du matériel d'usine moins énergivore. On pourrait mener une étude scientifique sur un échantillon de ce matériel afin de mesurer les économies d'énergie qu'il rendrait possibles. Dans une situation comme celle-là, il est manifestement indispensable d'avoir recours à des enquêteurs spécialisés capables de prendre les mesures nécessaires avec précision.

### **1.3 Instruments de mesure**

La collecte de données suppose généralement qu'on prenne des mesures. Or, comme la qualité d'une évaluation est fonction de celle des mesures prises, il faudrait prendre soin de se donner des instruments de mesure capables de produire des données valides et fiables. (Pour une excellente analyse de l'élaboration de questionnaires, voir Bradburn *et al.*, 1979.) Dans les enquêtes, l'instrument de mesure est un questionnaire; or, la préparation de questionnaires est loin d'être une science exacte. On estime d'ailleurs qu'au moins 20 à 30 p. 100 de la marge d'erreur des enquêtes est attribuable à l'ambiguïté des questions. Statistique Canada distribue à ce sujet un guide de sa conception intitulé *Conception d'un questionnaire de base*.

La conception d'un questionnaire comprend les cinq étapes suivantes :

### **Définir les concepts à mesurer**

Cela peut sembler étonnant, mais la tâche la plus difficile pour la conception d'un questionnaire consiste à préciser exactement l'information recherchée. À cette fin, il faut habituellement :

- dépouiller des études analogues, voire effectuer des recherches exploratoires;
- bien comprendre les questions à évaluer dans l'enquête;
- comprendre les concepts à mesurer et la façon de le faire;
- formuler l'hypothèse à évaluer;
- comprendre comment les réponses obtenues fourniront des indications sur les questions à évaluer;
- avoir une bonne idée du degré de validité et de fiabilité nécessaire pour arriver à produire de l'information, des données ou des éléments de preuve crédibles.

Avant de passer à l'étape suivante, la deuxième, il faut traduire les objectifs de la recherche en besoins d'information que l'enquête est susceptible de combler.

### **Choisir les questions (ou les points à mesurer) et établir les échelles de pondération**

Les questions peuvent être présentées de diverses façons (ouvertes ou fermées, à choix unique ou à choix multiples, et ainsi de suite). L'échelle choisie pour l'attribution de valeurs aux réponses éventuelles a elle aussi son importance, compte tenu de son incidence sur la validité des mesures.

### **Rédiger les questions**

C'est essentiellement un travail de communication, car il s'agit de savoir comment formuler des questions sans donner prise à l'ambiguïté ou à des biais, compte tenu des caractéristiques des répondants. Dans bien des secteurs de programme, il existe des questions et des mesures toutes faites dont l'évaluateur peut se servir avec profit. Par exemple, le Centre de recherche sur les enquêtes de l'Université du Michigan a décrit divers moyens de mesurer les attitudes psychosociologiques et évalué les avantages et les inconvénients de chacun d'entre eux (Robinson et Shaver, 1973).

## **Décider de l'enchaînement des questions et de la présentation du questionnaire**

Il faut que l'enchaînement des questions éveille l'intérêt des répondants, tout en ne provoquant aucun biais, comme celui qui se manifeste lorsque l'ordre des questions semble mener à une conclusion prédéterminée.

### **Faire un essai préalable du questionnaire**

Un essai préalable du questionnaire permet de détecter les questions ambiguës, les formulations boiteuses et les omissions. Cet essai devrait être réalisé auprès d'un petit échantillon de la population visée (voir Smith, 1975).

## **1.4 Estimation des coûts**

Pour estimer les coûts d'une enquête, il faut diviser son exécution en plusieurs éléments distincts, puis calculer le coût de revient de chacun d'entre eux, selon qu'ils seront réalisés à l'interne ou à l'externe. Le coût par entrevue pourrait être fondé sur les coûts de la conception de l'enquête, de la collecte et de la mise en forme des données, du codage et de la transcription des données brutes sous forme exploitable par machine, ainsi que de la compilation ou de l'analyse des données.

On peut confier des enquêtes à contrat aux groupes des enquêtes spéciales de Statistique Canada ou à des entreprises privées spécialisées. Statistique Canada publie d'ailleurs un répertoire des organismes d'enquête dans lequel leurs domaines de spécialisation sont précisés.

## **1.5 Avantages et inconvénients**

Nous allons maintenant passer à l'étude de trois méthodes d'enquête portant sur des personnes, dans le contexte des évaluations. Pour une analyse des avantages et des inconvénients des aspects statistiques des enquêtes, voir Smith (1975), le chapitre 8 et Galtung (1967).

### **Entrevues directes**

L'entrevue directe éveille l'intérêt des répondants et accroît le taux de participation. C'est une méthode qui permet à l'enquêteur de poser des questions complexes, pouvant exiger des explications ou des aides visuelles et mécaniques. Elle a aussi l'avantage de lui donner l'occasion de se faire préciser les réponses. On opte généralement pour elle lorsqu'il faut obtenir beaucoup de renseignements détaillés des répondants. En outre, elle est très souple, puisque l'enquêteur peut sauter les questions qui lui semblent non pertinentes et en poser d'autres. L'enquêteur peut aussi observer les caractéristiques des répondants et les noter. Qui plus est, c'est une méthode à laquelle on peut avoir recours lorsqu'il est impossible d'établir une base de sondage ou une liste des répondants. D'un autre côté, elle prend beaucoup de temps,

elle est difficile à administrer et à contrôler et, de plus, elle est très coûteuse. Enfin, elle est vulnérable aux biais attribuables à l'enquêteur et aux répondants loquaces, ce dernier se manifestant lorsque certaines personnes s'expriment plus ouvertement que d'autres, de sorte que leurs opinions sont plus en évidence.

### **Entrevues téléphoniques**

L'entrevue téléphonique est une méthode à la fois rapide, économique et facile à administrer et à contrôler, à condition d'être réalisée à partir d'un point central. Les résultats peuvent être entrés directement dans un ordinateur, si le système téléphonique est raccordé à un terminal, ce qui rend cette approche très efficace.

Ce genre d'entrevue est un excellent moyen d'avoir accès à des gens difficiles à joindre, comme des cadres supérieurs occupés. Par contre, lorsqu'on communique par téléphone, il est difficile de faire de longues entrevues, de poser des questions complexes ou d'utiliser les aides visuelles ou mécaniques. De plus, comme certaines personnes ont des numéros de téléphone confidentiels ou n'ont pas le téléphone, c'est une méthode qui peut comporter un biais attribuable à l'échantillonnage. Enfin, le biais attribuable à la non-réponse peut lui aussi poser un problème, puisque le répondant peut raccrocher n'importe quand. Il ne faut pas non plus oublier le risque de biais attribuables aux répondants loquaces.

### **Enquêtes postales**

Le principal avantage des enquêtes postales est leur coût modique; leur principal inconvénient est imputable au nombre élevé de variables dont il est impossible de tenir compte puisqu'il n'y a pas d'enquêteur, par exemple, l'identité du répondant, les personnes que celui-ci peut avoir consultées pour l'aider à répondre au questionnaire, la vitesse de réponse, l'ordre dans lequel les réponses sont données ou la compréhension qu'a le répondant des questions. Néanmoins, pour bien des types de questions, l'expérience a clairement prouvé que les enquêtes postales donnent des résultats plus précis que les autres méthodes d'enquête. De plus, elles permettent d'atteindre beaucoup de gens, et les répondants sont souvent plus ouverts lorsqu'ils répondent par écrit que lorsqu'ils doivent répondre de vive voix. Malheureusement, si cette méthode a l'avantage d'être peu coûteuse, elle a aussi l'inconvénient majeur d'un taux de réponse peu élevé et d'un biais attribuable à la non-réponse. En outre, les enquêtes postales exigent beaucoup de temps (pour l'envoi, le traitement et la réponse) et elles empêchent l'enquêteur d'approfondir et de clarifier certains points.

### **Résumé**

Comme nous l'avons vu, chaque méthode d'enquête a ses avantages et ses inconvénients. Pour l'évaluation, il faut tenir compte des facteurs suivants :

- exactitude (absence de biais);

- quantité de données à recueillir;
- souplesse (possibilité d'avoir recours à diverses techniques d'entrevue);
- biais attribuable à l'échantillonnage (possibilité de constituer un échantillon représentatif);
- biais attribuable à la non-réponse (risque que les non-répondants soient systématiquement différents des répondants);
- coût par entrevue;
- vitesse de réponse;
- faisabilité opérationnelle (possibilité de respecter les contraintes opérationnelles, telles que les coûts et le personnel).

Les enquêtes sur les objets impliquent la collecte de renseignements objectifs, habituellement plus valides et plus crédibles que les opinions et les impressions d'éventuels répondants. Pourtant, ces enquêtes ne sont pas exemptes de nombreuses erreurs, notamment d'échantillonnage (L'échantillon est-il bien représentatif des objets?) et de mesure (L'instrument de mesure utilisé est-il précis, et l'évaluateur s'en sert-il correctement?).

Enfin, si bien conçue soit-elle, l'enquête peut produire des données inutilisables lorsqu'elle est mal exécutée. Les enquêteurs doivent être bien formés. Il est essentiel de consacrer le temps et les ressources nécessaires à leur formation et à celle des préposés au codage. En effet, il est possible d'accroître la fiabilité et la validité des résultats en favorisant la plus grande uniformité possible de la compréhension du questionnaire qu'ont les enquêteurs et les codeurs, de leur compétence et des instructions qu'on leur donne.

**Notes :**



## Annexe 2

### GLOSSAIRE

**Ampleur** : Portée d'une mesure.

**Analyse coûts-avantages** : Analyse comparant les avantages que procure un programme aux coûts associés à son exécution. Une valeur monétaire est attribuée aux avantages et aux coûts.

**Analyse coût-efficacité** : Analyse comparant les coûts d'un programme à ses retombées. Dans cette analyse, les retombées ne sont pas traduites en valeur monétaire.

**Analyse coûts-avantages/coût-efficacité *ex ante*** : Analyse coûts-avantages ou coûts-efficacité portant non pas sur les avantages et les coûts réels d'un programme, mais sur des hypothèses de coûts et d'avantages établies *a priori*. Ce genre d'analyse est utilisé pour la planification plutôt que pour l'évaluation.

**Analyse coûts-avantages/coût-efficacité *ex post*** : Analyse coûts-avantages ou coûts-efficacité effectuée lorsqu'un programme fonctionne depuis un certain temps afin d'évaluer les coûts et les avantages *réels*.

**Analyse statistique** : Manipulation de données numériques ou catégoriques afin de prévoir des phénomènes, de tirer des conclusions sur des rapports entre variables ou de généraliser des résultats.

**Analyse statistique descriptive** : Chiffres et tableaux servant à résumer et à présenter succinctement une information quantitative.

**Analyse inférentielle statistique** : Analyse statistique utilisant des modèles pour confirmer les rapports entre variables ou pour généraliser les constatations à l'ensemble de la population.

**Appariement des sujets** : Division de la population en «blocs» établis selon une ou plusieurs variables autres que le programme susceptibles d'exercer une influence sur l'effet du programme.

**Aspects d'efficacité** : Catégorie d'aspects sur lesquels porte une évaluation, liés à la réalisation des objectifs d'un programme et aux autres conséquences et effets escomptés ou non du programme.

**Attribution** : Estimation de la mesure dans laquelle les résultats observés sont attribuables à un programme, ce qui signifie que le programme a eu des effets incrémentiels.

**Attrition** : Fait pour les participants à un traitement (ou les membres d'un groupe témoin) de délaissier le programme. Ce facteur peut nuire à la comparabilité des groupes expérimental et témoin et constituer un obstacle à la validité interne.

**Base de sondage** : Liste des éléments de la population sondée.

**Biais attribuable à la sélection** : Fait, pour les groupes expérimental et témoin relatifs à un programme, d'être au départ inégaux sur le plan statistique, pour un ou plusieurs facteurs importants. C'est un obstacle à la validité interne.

**Biais attribuable à l'enquêteur** : Influence que l'enquêteur exerce sur le répondant. Cette influence peut être attribuable à plusieurs facteurs, dont les caractéristiques physiques et psychologiques de l'enquêteur, qui peuvent susciter des réponses différentes selon le répondant.

**Biais attribuable à l'ordre des réponses** : Facteur de distorsion des résultats causé par l'ordre dans lequel les questions sont posées dans une enquête.

**Biais attribuable aux essais** : Changements observés dans le cadre d'une quasi-expérience qui peuvent être attribuables au fait que les participants connaissent bien l'instrument de mesure. C'est un obstacle possible à la validité interne.

**Biais attribuable aux instruments** : Conséquence d'un changement d'instrument selon la mesure lorsqu'on a recours à des enquêteurs différents. C'est un obstacle à la validité interne.

**Biais attribuable aux répondants loquaces** : Biais qui se produit lorsque certaines personnes s'expriment plus franchement que d'autres, et que leurs points de vue ressortent davantage.

**Biais de non-réponse** : Facteur de distorsion attribuable à la non-réponse : les réponses provenant d'unités d'échantillonnage qui fournissent une information peuvent ne pas correspondre aux réponses des unités d'échantillonnage qui ne répondent pas, et ce sur des aspects importants.

**Biais principaux** : Effets distincts de chaque variable expérimentale.

**Biais statistiquement significatif** : Biais observé et probablement pas exclusivement attribuable au hasard. Ce biais peut être vérifié au moyen de tests statistiques.

**Composition aléatoire** : Technique utilisée pour les entrevues par téléphone et permettant de choisir un échantillon. L'enquêteur compose un numéro à l'aide d'un système de composition aléatoire quelconque, sans savoir si ce numéro existe ni s'il s'agit du numéro d'une entreprise, d'un hôpital ou d'un ménage.

**Consultation de spécialistes :** Méthode de collecte des données faisant appel aux opinions et aux connaissances de spécialistes dans des domaines fonctionnels en tant qu'indicateurs des résultats d'un programme.

**Dépouillement de la documentation spécialisée :** Méthode de collecte des données qui comprend l'examen de rapports de recherche, de publications et de livres.

**Diffusion ou imitation du traitement :** Fait pour les répondants appartenant à un groupe témoin de ressentir eux aussi les effets destinés au groupe expérimental (exposé au programme). C'est un obstacle à la validité interne.

**Données longitudinales :** Données recueillies au cours d'une période; il peut aussi s'agir d'une série de données accumulées concernant des personnes ou des entités.

**Données objectives :** Observations dénuées d'impressions personnelles et fondées sur des faits observables. Les données objectives peuvent être mesurées quantitativement ou qualitativement.

**Données primaires :** Données recueillies par une équipe d'évaluation expressément pour l'évaluation.

**Données qualitatives :** Observations catégoriques plutôt que numériques portant souvent sur les attitudes, les perceptions et les intentions.

**Données quantitatives :** Observations numériques.

**Données secondaires :** Données recueillies et consignées par une autre personne ou une autre organisation (ordinairement à une date antérieure), habituellement à des fins autres que celles de l'évaluation en cours.

**Données subjectives :** Observations dans lesquelles entrent en jeu des sentiments, des attitudes et des perceptions personnelles. Les données subjectives peuvent être mesurées quantitativement ou qualitativement.

**Données transversales :** Données recueillies au même moment auprès de diverses entités.

**Écart type :** L'écart type d'un ensemble de mesures numériques (sur une «échelle d'intervalles») indique le degré de regroupement des mesures individuelles autour de la moyenne.

**Échantillonnage aléatoire :** Sélection d'unités d'une population fondée sur le principe de la répartition au hasard. Il existe pour chaque unité de la population une probabilité calculable (différente de zéro) d'être choisie.

**Échantillonnage non aléatoire :** Choix des unités d'un échantillon effectué de façon à ce que chaque unité de la population n'ait pas une probabilité calculable différente de zéro d'être choisie pour faire partie de l'échantillon.

**Échantillonnage par liste** : Technique principalement utilisée pour les entrevues téléphoniques afin de prélever un échantillonnage. L'enquêteur part d'une base de sondage renfermant des numéros de téléphone, choisit une unité dans cette base et réalise une entrevue téléphonique soit avec une personne précise, soit avec la première qui répond à ce numéro.

**Échantillonnage stratifié** : Technique d'échantillonnage aléatoire suivant laquelle une population est divisée en couches relativement homogènes appelées strates. Des échantillons appropriés sont choisis dans chaque strate.

**Échantillonnage subdivisé** : Technique d'échantillonnage aléatoire impliquant le choix d'un certain nombre d'échantillons indépendants à partir d'une population plutôt que celui d'un seul échantillon. Chaque sous-échantillon est appelé échantillon subdivisé et est choisi indépendamment des autres en fonction du même plan d'échantillonnage.

**Effet d'interaction** : Effet net combiné d'au moins deux variables qui influent sur le résultat d'une quasi-expérience.

**Enquête** : Méthode de collecte des données qui suppose une démarche planifiée en vue de recueillir les données requises auprès d'un échantillon de la population visée (ou au moyen d'un recensement complet). La population visée est composée des personnes ou des entités touchées par le programme (ou de personnes ou entités semblables).

**Entrevue non structurée** : Technique d'entrevue utilisant une conversation normale menant à des questions spontanées, souvent dans le cadre de l'observation régulière des activités d'un programme.

**Entrevue suivant une présentation type** : Technique d'entrevue utilisant des questions ouvertes et des questions fermées dont le texte est rédigé avant l'entrevue.

**Erreur autre que d'échantillonnage** : Type d'erreur non attribuable à l'échantillonnage se produisant dans presque toute activité d'enquête (même un recensement). Il peut s'agir par exemple de l'interprétation différente que les répondants donnent aux questions, d'erreurs de traitement des résultats ou d'erreurs dans la base de sondage.

**Erreur d'échantillonnage** : Erreur attribuée à l'échantillonnage et à la mesure d'un segment de la population pour éviter de devoir exécuter un recensement dans les mêmes conditions générales.

**Étude de cas** : Méthode de collecte des données qui suppose des études en profondeur de cas ou de projets liés à un programme. Cette méthode comporte une ou plusieurs techniques de collecte des données (p. ex., entrevues, étude de dossiers).

**Étude de dossiers** : Méthode de collecte des données impliquant l'examen des dossiers d'un programme. Il existe ordinairement deux sortes de dossiers : les dossiers à caractère général au sujet d'un programme et les dossiers portant sur des projets, clients ou participants particuliers.

**Événements historiques** : Événements non liés au programme, mais influant sur les réactions des intéressés.

**Exactitude** : Différence entre une estimation faite à partir d'un échantillon et des résultats obtenus à la suite d'un recensement. Dans les estimations non biaisées, précision et exactitude sont synonymes.

**Exhaustivité** : Étendue et profondeur maximales de l'examen des questions faisant l'objet de l'évaluation.

**Facteurs de régression** : Pseudo-changements des résultats d'un programme qui se produisent lorsqu'on a choisi pour un programme des personnes ou des unités de traitement en raison de leurs résultats extrêmes. Ces facteurs sont un obstacle à la validité interne.

**Fiabilité** : Degré auquel une mesure appliquée de façon répétée à une situation donnée produit les mêmes résultats, pourvu que la situation ne change pas d'une application à une autre. La fiabilité peut correspondre à la stabilité de la mesure dans le temps ou à la permanence de la mesure d'un endroit à un autre.

**Formule de la taille de l'échantillon** : Équation utilisée pour déterminer la taille minimale requise de l'échantillon. Cette équation varie selon la sorte d'estimation à faire, le degré de précision recherché et la méthode d'échantillonnage.

**Groupe de comparaison** : Groupe qui n'est pas exposé à un programme ou à un traitement. Voir également «groupe témoin».

**Groupe expérimental** : En recherche, groupe de sujets qui bénéficie du programme; aussi appelé groupe de traitement ou groupe exposé au programme.

**Groupe témoin** : Dans les modèles quasi-expérimentaux, groupe de sujets qui éprouve toutes les influences, sauf celles du programme, exactement de la même façon que le groupe exposé à un traitement (qu'on appelle aussi «groupe expérimental»). On peut l'appeler groupe non exposé au programme.

**Guide d'entrevue** : Liste de sujets à aborder ou de questions à poser au cours de l'entrevue.

**Hypothèses plausibles** : Autres façons possibles d'expliquer les résultats d'un programme, c'est-à-dire influences autres que celle du programme.

**Inférence causale** : Processus logique utilisé pour tirer des conclusions à partir des données ou des éléments de preuve sur les retombées d'un programme. Lorsqu'on dit qu'un programme a produit ou causé un certain résultat, cela signifie que, s'il n'avait pas existé (ou s'il avait existé sous une forme ou avec une ampleur différentes), le résultat obtenu (ou le niveau de résultat) ne se serait pas produit.

**Instruments de mesure** : Instruments utilisés pour recueillir des données (p. ex., questionnaires, directives d'entrevue, formulaires d'inscription des observations).

**Interaction entre la sélection et le programme** : Réceptivité inhabituelle des participants à un programme attribuable au fait qu'ils sont conscients de participer au programme ou à une enquête. C'est un obstacle à la validité interne et externe.

**Interaction entre le milieu et le programme** : Non-représentativité du milieu dans lequel se déroule le projet expérimental ou pilote par rapport au milieu envisagé pour le programme. C'est un obstacle à la validité externe.

**Interaction entre les événements historiques et le programme** : Conditions dans lesquelles s'est déroulé le programme et qui ne sont pas représentatives des conditions futures. C'est un obstacle à la validité externe.

**Maturation** : Changements des résultats attribuables au temps plutôt qu'au programme, par exemple le vieillissement des participants. C'est un obstacle à la validité interne.

**Méthode d'échantillonnage** : Méthode de choix des unités d'échantillonnage (p. ex., échantillonnage systématique, stratifié).

**Méthode de collecte des données** : Manière dont sont réunis les faits relatifs à un programme et à ses résultats. Le dépouillement de la documentation spécialisée, l'étude de dossiers, les observations directes, les enquêtes, la consultation de spécialistes et les études de cas figurent parmi les méthodes de collecte des données fréquemment utilisées pour l'évaluation de programmes.

**Méthodes analytiques directes** : Méthodes utilisées pour traiter les données afin de fournir des éléments de preuve sur les retombées ou les effets directs d'un programme.

**Modalités multiples d'établissement de la preuve** : Utilisation de plusieurs stratégies d'évaluation indépendantes pour examiner la même question d'évaluation, à partir de sources de données ou de méthodes analytiques des données différentes.

**Modèle d'entrées-sorties** : Modèle économique pouvant servir à analyser les dépendances mutuelles entre différents éléments d'une économie. C'est un modèle systématique qui présente les échanges de biens et de services entre les segments de production et de consommation d'une économie.

**Modèle d'évaluation** : Modèle logique ou cadre conceptuel utilisé pour tirer des conclusions au sujet des résultats.

**Modèle d'évaluation idéal** : Comparaison conceptuelle de deux ou de plusieurs situations identiques en tout point, sauf que le programme s'applique dans un seul cas. Un seul groupe (le groupe expérimental) bénéficie du programme; l'autre; les autres groupes (groupes témoins) sont exposés à toutes les influences pertinentes, sauf celles du programme, exactement de la même façon que le groupe expérimental. Les résultats sont mesurés de manière identique pour chaque groupe, et toute différence observée peut être attribuée au programme.

**Modèle implicite** : Modèle dans lequel il n'y a pas de groupe témoin officiel, et où les mesures sont prises après l'exposition au programme.

**Modèle macro-économique** : Modèle des interactions entre les marchés des produits, de la main-d'œuvre et des biens d'une économie qui s'intéresse aux niveaux de production et des prix, compte tenu des interactions entre l'offre et la demande globales.

**Modèle micro-économique** : Modèle du comportement économique des acheteurs et vendeurs individuels sur un marché donné et dans des circonstances particulières.

**Modèle quasi expérimental** : Structure d'étude utilisant des groupes de comparaison pour faire des inférences causales, mais sans recours à la randomisation pour constituer un groupe expérimental et un groupe témoin. Le premier groupe est ordinairement acquis; le groupe expérimental est choisi pour correspondre le plus possible avec lui, de façon à permettre des inférences sur les effets incrémentiels du programme.

**Modèle statistique** : Modèle ordinairement fondé sur des recherches antérieures et permettant de transformer une mesure précise des effets en une autre mesure précise des effets, une mesure précise des effets en une gamme d'autres mesures des effets ou une gamme de mesures des effets en une autre gamme de mesures des effets.

**Modèles expérimentaux (ou aléatoires)** : Modèles utilisés pour établir l'équivalence initiale entre un ou plusieurs groupes témoins et le groupe exposé à un traitement en créant administrativement des groupes par assignation aléatoire, ce qui permet d'en assurer l'équivalence mathématique. Comme exemples de modèles expérimentaux ou aléatoires, signalons les suivants : les modèles avec groupes aléatoires, les modèles à carré latin, les modèles fractionnels et les quatre groupes de Salomon.

**Niveau de confiance** : Affirmation selon laquelle la valeur réelle d'un paramètre pour une population donnée se situe à l'intérieur d'un certain niveau de probabilité dans une fourchette spécifiée de valeurs.

**Non-réponse** : Situation qui se produit lorsqu'on ne peut pas obtenir d'information auprès d'unités d'échantillonnage.

**Objectivité** : Qualité des éléments de preuve et des conclusions qui peut être vérifiée par une personne autre que les auteurs.

**Observation sur le terrain** : Méthode de collecte des données qui suppose des visites aux endroits où un programme est exécuté. Elle a pour but d'évaluer directement le contexte du programme, ses activités et les personnes qui y participent.

**Politique monétaire** : Mesure de l'État exerçant une influence sur la masse monétaire et les taux d'intérêt. Il peut aussi s'agir d'un programme.

**Population** : Ensemble des unités auxquelles s'appliquent les résultats d'une enquête.

**Profondeur** : Degré d'exactitude et de détail d'une mesure.

**Randomisation** : Utilisation d'un plan de probabilité pour établir un échantillon. On peut utiliser des tables de nombres aléatoires, des ordinateurs, des dés, des cartes, etc.

**Stratégie d'évaluation** : Méthode utilisée pour recueillir des données sur les retombées d'un programme. Elle comprend un modèle d'évaluation, une méthode de collecte des données et une technique d'analyse.

**Taille de l'échantillon** : Nombre d'unités à échantillonner.

**Unité d'échantillonnage** : Unité utilisée pour l'échantillonnage. La population devrait être divisible en un nombre fini d'unités distinctes qui ne se chevauchent pas, afin que chaque membre de la population n'appartienne qu'à une unité d'échantillonnage.

**Validité de la mesure** : Une mesure est valable si elle représente ce qu'elle est censée représenter. Les mesures valables ne présentent pas de biais systématique.

**Validité des conclusions** : Aptitude à généraliser les conclusions tirées d'un programme actuel et à les appliquer à d'autres endroits, lieux ou situations. Pour formuler des conclusions généralisables, il faut satisfaire aux critères de validité interne et de validité externe.

**Validité externe** : Aptitude à généraliser les conclusions tirées d'un programme et à les appliquer à des conditions futures ou différentes. Les obstacles à la validité externe comprennent l'interaction entre la sélection et le programme, l'interaction entre le milieu et le programme et l'interaction entre les événements historiques et le programme.

**Validité interne** : Aptitude à affirmer qu'un programme a eu des résultats mesurés (jusqu'à un certain point), malgré d'autres explications plausibles. Les obstacles à la validité interne les plus courants sont les événements historiques, la maturation, l'attrition, les biais attribuables à la sélection, les facteurs de régression statistique, la diffusion et l'imitation du traitement, ainsi que l'essai.



## Annexe 3

### BIBLIOGRAPHIE

- Abt, C.G., éd., *The Evaluation of Social Programs*, Thousand Oaks : Sage Publications, 1976.
- Alberta, ministère du Trésor, *Measuring Performance : A Reference Guide*, septembre 1996.
- Alkin, M.C., *A Guide for Evaluation Decision Makers*, Thousand Oaks : Sage Publications, 1986.
- Angelsen, Arild et Ussif Rashid Sumaila, *Hard Methods for Soft Policies : Environmental and Social Cost-Benefit Analysis*, Bergen, Norvège : Institut Michelsen, 1995.
- Australie, ministère des Finances, *Handbook of Cost-Benefit Analysis*, Canberra, 1991.
- Babbie, E.R., *Survey Research Methods*, Belmont : Wadsworth, 1973.
- Baird, B.F., *Managerial Decisions Under Uncertainty*, New York : Wiley Interscience, 1989.
- Behn, R.D. et J.W. Vaupel, *Quick Analysis for Busy Division Makers*, New York : Basic Books, 1982.
- Belli, P., *Guide to Economic Appraisal of Development Projects*, Washington (DC) : Banque mondiale, 1996.
- Bentkover, J.D., Covdlo, V.T. et J. Mumpower, *Benefits Assessment : The State of the Art*, Dordrecht, Pays-Bas : D. Reidel Publishing Co., 1986.
- Berk, Richard A. et Peter H. Rossi, *Thinking About Program Evaluation*, Thousand Oaks : Sage Publications, 1990.
- Bickman, L., éd., *Using Program Theory in Program Evaluation*, Vol. 33 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1987.
- Blalock, H.M., Jr., *Measurement in the Social Sciences : Theories and Strategies*, Chicago : Aldine, 1974.

Blalock, H.M., Jr., éd., *Causal Models in the Social Sciences*, Chicago : Aldine, 1971.

Boberg, Alice L. et Sheryl A. Morris-Khoo, «The Delphi Method : A Review of Methodology and an Application in the Evaluation of a Higher Education Program», *Canadian Journal of Program Evaluation*, Vol. 7, N° 1, avril-mai, 1992, p. 27 à 40.

Boruch, R.F., «Conducting Social Experiments», *Evaluation Practice in Review*, Vol. 34 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1987, p. 45 à 66.

Boruch, R.F., et al., *Reanalysing Program Evaluations — Policies and Practices for Secondary Analysis for Social and Education Programs*, San Francisco : Jossey-Bass, 1981.

Boruch, R.F., «On Common Contentions About Randomized Field Experiments», in Glass, Gene V., éd., *Evaluation Studies Review Annual*, Thousand Oaks : Sage Publications, 1976.

Bradburn, N.M. et S. Sudman, *Improving Interview Methods and Questionnaire Design*, San Francisco : Jossey-Bass, 1979.

Braverman, Mark T. et Jana Kay Slater, *Advances in Survey Research*, Vol. 70 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1996.

Buffa, E.S. et J.S. Dyer, *Management Science Operations Research : Model Formulation and Solution Methods*, New York : John Wiley and Sons, 1977.

Cabatoff, Kenneth A., «Getting On and Off the Policy Agenda : A Dualistic Theory of Program Evaluation Utilization», *Canadian Journal of Program Evaluation*, Vol. 11, N° 2, automne 1996, p. 35 à 60.

Campbell, D., «Considering the Case Against Experimental Evaluations of Social Innovations», *Administrative Science Quarterly*, Vol. 15, N° 1, 1970, p. 111 à 122.

Campbell, D.T., «Degrees of Freedom and the Case Study», *Comparative Political Studies*, Vol. 8, 1975, p. 178 à 193.

Campbell, D.T. et J.C. Stanley, *Experimental and Quasi-Experimental Designs for Research*, Chicago : Rand-McNally, 1963.

Canada, Bureau du Vérificateur général du Canada, *Bulletin 84-7, Photographies et autres aides visuelles*.

Canada, Bureau du Vérificateur général, «Le choix et l'application des techniques de collecte des éléments probants en vérification d'optimisation des ressources», *Analyse coûts-avantages*, 1994, annexe B5.

Canada, Secrétariat du Conseil du Trésor, *Aborder les années 90 : Perspectives gouvernementales pour l'évaluation de programmes*, Ottawa, 1991.

Canada, Secrétariat du Conseil du Trésor, «Examen, vérification interne et évaluation», *Manuel du Conseil du Trésor*, Ottawa, dernière révision en 1994.

Canada, Secrétariat du Conseil du Trésor, *Guide de l'analyse avantages-coûts*, Ottawa, 1997.

Canada, Secrétariat du Conseil du Trésor, *Guide de la gestion de la qualité*, Ottawa, octobre 1992.

Canada, Secrétariat du Conseil du Trésor, *Guides des services de qualité :*

*Services de qualité – Tour d'horizon*, Ottawa, octobre 1995.

*Guide I – Consultation des clients*, Ottawa, octobre 1995.

*Guide II – Mesure de la satisfaction des clients*, Ottawa, octobre 1995.

*Guide III – Collaboration avec les syndicats*, Ottawa, octobre 1995.

*Guide IV – Un milieu propice à l'apprentissage*, Ottawa, octobre 1995.

*Guide V – Reconnaissance du mérite*, Ottawa, octobre 1995.

*Guide VI – Sondage auprès des employés*, Ottawa, octobre 1995.

*Guide VII – Normes de service*, octobre 1995.

*Guide VIII – Analyses comparatives et meilleures pratiques*, Ottawa, octobre 1995.

*Guide IX – Communications*, Ottawa, octobre 1995.

*Guide X – Analyse comparative et partage des pratiques exemplaires – Mise à jour du*

*Guide XI – Gestion efficace des plaintes*, Ottawa, juin 1996.

*Guide XII – Qui est le client? – Document de travail*, Ottawa, juillet 1996.

*Guide XIII – Guide des gestionnaires pour la prestation de services de qualité*, Ottawa, septembre 1996.

Canada, Secrétariat du Conseil du Trésor, *L'évaluation des programmes fédéraux : Répertoire sur l'utilisation des évaluations*, Ottawa, 1991.

Canada, Secrétariat du Conseil du Trésor, *Les normes de service : Un guide pour l'initiative*, Ottawa, février 1995.

Canada, Secrétariat du Conseil du Trésor, *Mesure de la satisfaction des clients – Concevoir et adopter de saines pratiques de mesure et de suivi de la satisfaction des clients*, Ottawa, octobre 1991.

Canada, Secrétariat du Conseil du Trésor, *Normes d'évaluation de programmes dans les ministères et organismes fédéraux*, Ottawa, juillet 1989.

Canada, Secrétariat du Conseil du Trésor, *Pour offrir aux Canadiens et aux Canadiennes un service de qualité à un prix raisonnable – Établissement de normes de service au gouvernement fédéral*, Ottawa, décembre 1994.

Canada, Secrétariat du Conseil du Trésor, *Pour une fonction d'examen plus efficace – Rapport annuel au Parlement par le Président du Conseil du Trésor*, Ottawa, octobre 1995.

Canada, Secrétariat du Conseil du Trésor, *Repenser le rôle de l'État : Améliorer la mesure des résultats et de la responsabilisation – Rapport annuel au Parlement par le Président du Conseil du Trésor*, Ottawa, octobre 1996.

Canada, Statistique Canada, *La structure par entrées-sorties de l'économie canadienne 1961-1981*, Ottawa, avril 1989, n° de cat. 15-201F.

Canada, Statistique Canada, *Lignes directrices concernant la qualité* (2<sup>e</sup> édition), Ottawa, 1987.

Canada, Statistique Canada, *Répertoire des méthodes d'évaluation des erreurs dans les recensements et les enquêtes*, Ottawa, 1982, CSCCB-F.

Caron, Daniel J., «Knowledge Required to Perform the Duties of an Evaluator», *Canadian Journal of Program Evaluation*, Vol. 8, N° 1, avril-mai 1993, p. 59 à 78.

Casley, D.J. et K. Kumar, *The Collection, Analysis and Use of Monitoring and Evaluation Data*, Washington (DC) : Banque mondiale, 1989.

Chatterjee, S. et B. Price, *Regression Analysis by Example* (2<sup>e</sup> édition), New York : John Wiley and Sons, 1995.

Chelimsky, Eleanor; éd., *Program Evaluation : Patterns and Directions*, Washington : American Society for Public Administration, 1985.

Chelimsky, Eleanor et William R. Shadish, éd., *Evaluation for the 21st Century : A Handbook*, Thousand Oaks : Sage Publications, 1997.

Chen, H.T. et P.H. Rossi, «Evaluating with Sense : the Theory-Driven Approach», *Evaluation Review*, Vol. 7, 1983, p. 283 à 302.

Chen, Huey-Tsyh, *Theory-Driven Evaluations*, Thousand Oaks : Sage Publications, 1990.

Chenery, H. et P. Clark, *Inter-industry Economics*, New York : John Wiley and Sons, 1959.

Ciarlo, J., éd., *Utilizing Evaluation*, Thousand Oaks : Sage Publications, 1984.

Clemen, R.T., *Making Hard Decisions*, Duxbury Press, 1991, sections 1 à 3.

Cook, T.D. et D.T. Campbell, *Quasi-Experimentation : Designs and Analysis Issues for Field Settings*, Chicago : Rand-McNally, 1979.

Cook, T.D. et C.S. Reichardt, éd., *Qualitative and Quantitative Methods in Evaluation Research*, Thousand Oaks : Sage Publications, 1979.

Cordray, D.S., «Quasi-Experimental Analysis : A Mixture of Methods and Judgement», in Trochim, W.M.K., éd., *Advances in Quasi-Experimental Design and Analysis*, p. 9 à 27, Vol. 31 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1986.

Datta L. et R. Perloff., *Improving Evaluations*, Thousand Oaks : Sage Publications, 1979, section II.

Delbecq, A.L., et al., *Group Techniques in Program Planning : A Guide to the Nominal Group and Delphi Processes*, Glenview : Scott, Foresman, 1975.

Dexter, L.A., *Elite and Specialized Interviewing*, Evanston (Illinois) : Northwestern University Press, 1970.

Duncan, B.D., *Introduction to Structural Equation Models*, New York : Academic Press, 1975.

Eaton, Frank, «Measuring Program Effects in the Presence of Selection Bias : The Evolution of Practice», *Canadian Journal of Program Evaluation*, Vol. 9, N° 2, octobre-novembre 1994, p. 57 à 70.

Favaro, Paul et Marie Billinger, «A Comprehensive Evaluation Model for Organizational Development», *Canadian Journal of Program Evaluation*, Vol. 8, N° 2, octobre-novembre 1993, p. 45 à 60.

Fienberg, S., *The Analysis of Cross-classified Categorical Data* (2<sup>e</sup> édition), Cambridge : MIT (Massachusetts Institute of Technology), 1980.

Fitzgibbon, C.T. et L.L. Morris, *Evaluator's Kit* (2<sup>e</sup> édition), Thousand Oaks : Sage Publications, 1988.

Fowler, Floyd J., *Improving Survey Questions : Design and Evaluation*, Thousand Oaks : Sage Publications, 1995.

Fox, J., *Linear Statistical Models and Related Methods, with Applications to Social Research*, New York : Wiley, 1984.

Gauthier, B., éd., *Recherche sociale : de la problématique à la collecte des données*, Montréal : Les Presses de l'Université du Québec, 1984.

Gliksman, Louis, et al., «Responders vs. Non-Responders to a Mail Survey : Are They Different?», *Canadian Journal of Program Evaluation*, Vol. 7, N° 2, octobre-novembre 1992, p. 131 à 138.

- Globerson, Aryé, *et al.*, *You Can't Manage What You Don't Measure : Control and Evaluation in Organizations*, Brookfield : Gower Publications, 1991.
- Goldberger, A.S. et D.D. Duncan, *Structural Equation Models in the Social Sciences*, New York : Seminar Press, 1973.
- Goldman, Francis et Edith Brashares, «Performance and Accountability : Budget Reform in New Zealand», *Public Budgeting and Finance*, Vol. 11, N° 4, hiver 1991, p. 75 à 85.
- Goode, W.J. et Paul K. Hutt, *Methods in Social Research*, New York : McGraw-Hill, 1952, chapitre 9.
- Gordon, R.A., *Economic Instability and Growth : The American Record*, Harper & Row, 1974.
- Guba, E.G., «Naturalistic Evaluation», in Cordray, D.S., *et al.*, éd., *Evaluation Practice in Review*, Vol. 34 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1987.
- Guba, E.G. et Y.S. Lincoln, *Effective Evaluation : Improving the Usefulness of Evaluation Results Through Responsive and Naturalistic Approaches*, San Francisco : Jossey-Bass, 1981.
- Hanley, J.A., «Appropriate Uses of Multivariate Analysis», *Annual Review of Public Health*, Palo Alto : Annual Reviews Inc., 1983, p. 155 à 180.
- Hanushek, E.A. et J.E. Jackson, *Statistical Methods for Social Scientists*, New York : Academic Press, 1977.
- Harberger, A.C., *Project Evaluation : Collected Papers*, Chicago : Markham Publishing Co., 1973.
- Heilbroner, R.L. et L.C. Thurow, *Economics Explained*, Toronto : Simon and Schuster Inc., 1987.
- Heise, D.R., *Causal Analysis*, New York : Wiley, 1975.
- Henderson, J. et R. Quandt, *Micro-economic Theory*, New York : McGraw-Hill, 1961.
- Hoaglin, D.C., *et al.*, *Data for Decisions*, Cambridge (MA) : Abt Books, 1982.
- Hudson, Joe, *et al.*, éd., *Action Oriented Evaluation in Organizations : Canadian Practices*, Toronto : Wall and Emerson, 1992.
- Huff, D., *How to Lie with Statistics*, Penguin, 1973.

- Jolliffe, R.F., *Common Sense Statistics for Economists and Others*, Routledge and Kegan Paul, 1974.
- Jorjani, Hamid, «The Holistic Perspective in the Evaluation of Public Programs : A Conceptual Framework», *Canadian Journal of Program Evaluation*, Vol. 9, N° 2, octobre-novembre 1994, p. 71 à 92.
- Katz, W.A., *Introduction to Reference Work : Reference Services and Reference Processes*, Vol. II, New York : McGraw-Hill, 1982, chapitre 4.
- Kenny, D.A., *Correlation and Causality*, Toronto : John Wiley and Sons, 1979.
- Kerlinger, F.N., *Behavioural Research : A Conceptual Approach*, New York : Holt, Rinehart and Winston, 1979.
- Kidder, L.H. et M. Fine, «Qualitative and Quantitative Methods : when Stories Converge», in *Multiple Methods in Program Evaluation*, Vol. 35 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1987.
- Kish, L., *Survey Sampling*, New York : Wiley, 1965.
- Krause, Daniel Robert, *Effective Program Evaluation : An Introduction*, Chicago : Nelson-Hall, 1996.
- Krueger, R.A., *Focus Groups : A Practical Guide for Applied Research*, Thousand Oaks : Sage Publications, 1988.
- Leeuw, Frans L., «Performance Auditing and Policy Evaluation : Discussing Similarities and Dissimilarities», *Canadian Journal of Program Evaluation*, Vol. 7, N° 1, avril-mai 1992, p. 53 à 68.
- Leontief, W., *Input-Output Economics*, New York : Oxford University Press, 1966.
- Levine, M., «Investigative Reporting as a Research Method : an Analysis of Bernstein and Woodward's 'All The President's Men'», *American Psychologist*, Vol. 35, 1980, p. 626 à 638.
- Love, Arnold J., *Evaluation Methods Sourcebook II*, Ottawa : Société canadienne d'évaluation, 1995.
- Mark, M.M., «Validity Typologies and the Logic and Practice of Quasi-Experimentation», in W.M.K. Trochim, éd., *Advances in Quasi-Experimental Design and Analysis*, p. 47 à 66, Vol. 31 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1986.
- Martin, Lawrence L. et Peter M. Kettner, *Measuring the Performance of Human Service Programs*, Thousand Oaks : Sage Publications, 1996.

- Martin, Michael O. et V.S. Mullis, éd., *Quality Assurance In Data Collection*, Chestnut Hill : Center for the Study of Testing, Evaluation, and Educational Policy, Boston College, 1996.
- Maxwell, Joseph A., *Qualitative Research Design : An Interactive Approach*, Thousand Oaks : Sage Publications, 1996.
- Mayne, John et Eduardo Zapico-Goñi, *Monitoring Performance in the Public Sector : Future Directions From International Experience*, New Brunswick (NJ) : Transaction Publishers, 1996.
- Mayne, John, et al., éd., *Advancing Public Policy Evaluation : Learning From International Experiences*, Amsterdam : North-Holland, 1992.
- Mayne, John et R.S. Mayne, «Will Program Evaluation be Used in Formulating Policy?», in Atkinson, M. et M. Chandler, éd., *The Politics of Canadian Public Policy*, Toronto : University of Toronto Press, 1983.
- Mayne, John, «In Defense of Program Evaluation», *Canadian Journal of Program Evaluation*, Vol. 1, N° 2, 1986, p. 97 à 102.
- McClintock, C.C., et al., «Applying the Logic of Sample Surveys to Qualitative Case Studies : The Case Cluster Method», in Van Maanen, J., éd., *Qualitative Methodology*, Thousand Oaks : Sage Publications, 1979.
- Mercer, Shawna L. et Vivek Goel, «Program Evaluation in the Absence of Goals : A Comprehensive Approach to the Evaluation of a Population-Based Breast Cancer Screening Program», *Canadian Journal of Program Evaluation*, Vol. 9, N° 1, avril-mai 1994, p. 97 à 112.
- Miles, M.B. et A.M. Huberman, *Qualitative Data Analysis : A Sourcebook and New Methods*, Thousand Oaks : Sage Publications, 1984.
- Miller, J.C. III et B. Yandle, *Benefit-Cost Analyses of Social Regulation*, Washington : American Enterprise Institute, 1979.
- Moore, M.H., *Creating Public Value : Strategic Management in Government*, Boston : Harvard University Press, 1995.
- Morris, C.N. et J.E. Rolph, *Introduction to Data Analysis and Statistical Inference*, Englewood Cliffs : Prentice Hall, 1981.
- Mueller, J.H., *Statistical Reasoning in Sociology*, Boston : Houghton Mifflin, 1977.
- Nachmias, C. et D. Nachmias, *Research Methods in the Social Sciences*, New York : St. Martin's Press, 1981, chapitre 7.



- Nelson, R., Merton, P. et E. Kalachek, *Technology, Economic Growth and Public Policy*, Washington (DC) : Brookings Institute, 1967.
- Nutt, P.C. et R.W. Backoff, *Strategic Management of Public and Third Sector Organizations*, San Francisco : Jossey-Bass, 1992.
- O'Brecht, Michael, «Stakeholder Pressures and Organizational Structure», *Canadian Journal of Program Evaluation*, Vol. 7, N° 2, octobre-novembre 1992, p. 139 à 147.
- Okun, A., *The Political Economy of Prosperity*, Norton, 1970.
- Paquet, Gilles et Robert Shepherd, «The Program Review Process : A Deconstruction», Ottawa : Faculté d'administration, Université d'Ottawa, 1996.
- Patton, M.Q., *Qualitative Evaluation Methods*, Thousand Oaks : Sage Publications, 1980.
- Patton, M.Q., *Creative Evaluation* (2<sup>e</sup> édition), Thousand Oaks : Sage Publications, 1986.
- Patton, M.Q., *Practical Evaluation*, Thousand Oaks : Sage Publications, 1982.
- Patton, M.Q., *Utilization-Focused Evaluation* (2<sup>e</sup> édition), Thousand Oaks : Sage Publications, 1986.
- Pearsol, J.A., éd., «Justifying Conclusions in Naturalistic Evaluations», *Evaluation and Program Planning*, Vol. 10, N° 4, 1987, p. 307 à 358.
- Perret, Bernard, «Le contexte français de l'évaluation : Approche comparative», *Canadian Journal of Program Evaluation*, Vol. 9, N° 2, octobre-novembre 1994, p. 93 à 114.
- Peters, Guy B. et Donald J. Savoie, *Governance in a Changing Environment*, Centre canadien de gestion, Montréal & Kingston : McGill-Queen's University Press, 1993.
- Polkinghorn, R.S., *Micro-theory and Economic Choices*, Richard Irwin Inc., 1979.
- Posavac, Emil J. et Raymond G. Carey, *Program Evaluation : Methods and Case Studies* (5<sup>e</sup> édition), Upper Saddle River (NJ) : Prentice Hall, 1997.
- Pressman, J.L. et A. Wildavsky, *Implementation*, Los Angeles : UCLA Press, 1973.
- Ragsdale, C.T., *Spreadsheet Modelling and Decision Analysis*, Cambridge : Course Technology Inc., 1995.
- Reavy, Pat, *et al.*, «Evaluation as Management Support : The Role of the Evaluator», *Canadian Journal of Program Evaluation*, Vol. 8, N° 2, octobre-novembre 1993, p. 95 à 104.

Rindskopf, D., «New Developments in Selection Modeling for Quasi-Experimentation», in W.M.K. Trochim, éd., *Advances in Quasi-Experimental Design and Analysis*, p. 79 à 89, Vol. 31 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1986.

Rist, Ray C., éd., *Program Evaluation and the Management of the Government*, New Brunswick (NJ) : Transaction Publishers, 1990.

Robinson, J.P. et P.R. Shaver, *Measurement of Social Psychological Attitudes*, Ann Arbor : Survey Research Centre, University of Michigan, 1973.

Rossi, P.H. et H.E. Freeman, *Evaluation : A Systematic Approach* (2<sup>e</sup> édition), Thousand Oaks : Sage Publications, 1989.

Rossi, P.H., Wright, J.D. et A.B. Anderson, éd., *Handbook of Survey Research*, Orlando : Academic Press, 1985.

Rush, Brian et Alan Ogborne, «Program Logic Models : Expanding their Role and Structure for Program Planning and Evaluation», *Canadian Journal of Program Evaluation*, Vol. 6, N<sup>o</sup> 2, octobre-novembre 1991, p. 95 à 106.

Rutman, L. et John Mayne, «Institutionalization of Program Evaluation in Canada : The Federal Level», in M.Q. Patton, éd., *Culture and Evaluation*, Vol. 25 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1985.

Ryan, Allan G. et Caroline Krentz, «All Pulling Together : Working Toward a Successful Evaluation», *Canadian Journal of Program Evaluation*, Vol. 9, N<sup>o</sup> 2, octobre-novembre 1994, p. 131 à 150.

Ryan, Brenda et Elizabeth Townsend, «Criteria Mapping», *Canadian Journal of Program Evaluation*, Vol. 4, N<sup>o</sup> 2, octobre-novembre 1989, p. 47 à 58.

Samuelson, P., *Foundations of Economic Analysis*, Cambridge : Harvard University Press, 1947.

Sang, H.K., *Project Evaluation*, New York : Wilson Press, 1988.

Sassone, P.G. et W.A. Schaffer, *Cost-Benefit Analysis : A Handbook*, New York : Academic Press, 1978.

Schick, Allen, *The Spirit of Reform : Managing the New Zealand State*, rapport commandé par le ministère du Trésor et la Commission des services aux États de la Nouvelle-Zélande, 1996.

Schmid, A.A., *Benefit-Cost Analysis : A Political Economy Approach*, Boulder : Westview Press, 1989.

- Seidle, Leslie., *Rethinking the Delivery of Public Services to Citizens*, Montréal : Institut de recherche en politiques publiques (IRPP), 1995.
- Self, P., *Econocrats and the Policy Process : The Politics and Philosophy of Cost-Benefit Analysis*, Londres : MacMillan, 1975.
- Shadish, William R., et al. *Foundations of Program Evaluation : Theories of Practice*, Thousand Oaks : Sage Publications, 1991.
- Shea, Michael P. et John H. Lewko, «Use of a Stakeholder Advisory Group to Facilitate the Utilization of Evaluation Results», *Canadian Journal of Program Evaluation*, Vol. 10, N° 1, avril-mai 1995, p. 159 à 162.
- Shea, Michael P. et Shelagh M.J. Towson, «Extent of Evaluation Activity and Evaluation Utilization of CES Members», *Canadian Journal of Program Evaluation*, Vol. 8, N° 1, avril-mai 1993, p. 79 à 88.
- Silk, L., *The Economists*, Avon Books, 1976.
- Simon, H., «Causation», in D.L. Sill, éd., *International Encyclopedia of the Social Sciences*, Vol. 2, New York : Macmillan, 1968, p. 350 à 355.
- Skaburskis, Andrejs et Fredrick C. Collignon, «Cost-Effectiveness Analysis of Vocational Rehabilitation Services», *Canadian Journal of Program Evaluation*, Vol. 6, N° 2, octobre-novembre 1991, p. 1 à 24.
- Skelton, Ian, «Sensitivity Analysis in Multi-criteria Decision Aids : A Demonstration of Child Care Need Assessment», *Canadian Journal of Program Evaluation*, Vol. 8, N° 1, avril-mai 1993, p. 103 à 116.
- Société canadienne d'évaluation, Comité de normalisation, «Standards for Program Evaluation in Canada : A Discussion Paper», *Canadian Journal of Program Evaluation*, Vol. 7, N° 1, avril-mai 1992, p. 157 à 170.
- Sprent, P., *Statistics in Action*, Penguin, 1977.
- Stolzenberg, J.R.M. et K.C. Land, «Causal Modeling and Survey Research», in Rossi, P.H., et al., *TITRE MANQUANT*, Orlando : Academic Press, 1983, p. 613 à 675.
- Stouthamer-Loeber, Magda et Welmoet Bok van Kammen, *Data Collection and Management : A Practical Guide*, Thousand Oaks : Sage Publications, 1995.
- Suchman, E.A., *Evaluative Research : Principles and Practice in Public Service and Social Action Programs*, New York : Russell Sage, 1967.
- Sugden, R. et A. Williams, *The Principles of Practical Cost-benefit Analysis*, Oxford : Oxford University Press, 1978.

- Tellier, Luc-Normand, *Méthodes d'évaluation des projets publics*, Sainte-Foy : Presses de l'Université du Québec, 1994, 1995.
- Thomas, Paul G., «The Politics and Management of Performance Measurement and Service Standards», Winnipeg : St-John's College, University of Manitoba, 1996.
- Thompson, M., *Benefit-Cost Analysis for Program Evaluation*, Thousand Oaks : Sage Publications, 1980.
- Thurston, W.E., «Decision-Making Theory and the Evaluator», *Canadian Journal of Program Evaluation*, Vol. 5, N° 2, octobre-novembre 1990, p. 29 à 46.
- Trochim, W.M.K., éd., *Advances in Quasi-Experimental Design and Analysis*, Vol. 31 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1986.
- Uhl, Norman et Carolyn Wentzel, «Evaluating a Three-Day Exercise to Obtain Convergence of Opinion», *Canadian Journal of Program Evaluation*, Vol. 10, N° 1, avril-mai 1995, p. 151 à 158.
- Van Pelt, M. et R. Timmer, *Cost-Benefit Analysis for Non-Economists*, Institut d'économie des Pays-Bas, 1992.
- Van Maasen, J., éd., *Qualitative Methodology*, Thousand Oaks : Sage Publications, 1983.
- Warwick, D.P. et C.A. Lininger, *The Survey Sample : Theory and Practice*, New York : McGraw-Hill, 1975.
- Watson, D.S., *Price Theory in Action*, Boston : Houghton Mifflin, 1970.
- Watson, Kenneth, «Selecting and Ranking Issues in Program Evaluations and Value-for-Money Audits», *Canadian Journal of Program Evaluation*, Vol. 5, N° 2, octobre-novembre 1990, p. 15 à 28.
- Watson, Kenneth, «The Social Discount Rate», *Canadian Journal of Program Evaluation*, Vol. 7, N° 1, avril-mai 1992, p. 99 à 118.
- Webb, E.J., *et al.*, *Nonreactive Measures in the Social Sciences* (2<sup>e</sup> édition), Boston : Houghton Mifflin, 1981.
- Weisberg, Herbert F., Krosnick, Jon A. et Bruce D. Bowen, éd., *An Introduction to Survey Research, Polling, and Data Analysis*, Thousand Oaks : Sage Publications, 1996.
- Weisler, Carl E., U.S. General Accounting Office, *Review Topics in Evaluation : What Do You Mean by Secondary Analysis?*
- Williams, D.D., éd., *Naturalistic Evaluation*, Vol. 30 de *New Directions in Program Evaluation*, San Francisco : Jossey-Bass, 1986.

Wye, Christopher G. et Richard C. Sonnichsen, éd., *Evaluation in the Federal Government : Changes, Trends and Opportunities*, San Francisco : Jossey-Bass, 1992.

Yates, Brian T., *Analyzing Costs, Procedures, Processes, and Outcomes in Human Services*, Thousand Oaks : Sage Publications, 1996.

Yin, R., *The Case Study as a Rigorous Research Method*, Thousand Oaks : Sage Publications, 1986.

Zanakis, S.H., *et al.*, «A Review of Program Evaluation and Fund Allocation Methods within the Service and Government», *Socio-Economic Planning Sciences*, Vol. 29, N° 1, mars 1995, p. 59 à 79.

Zúñiga, Ricardo, *L'évaluation dans l'action : choix de buts et choix de procédures*, Montréal : Librairie de l'Université de Montréal, 1992.

## Annexe 4

### AUTRES RÉFÉRENCES

*Administrative Science Quarterly*

*American Sociological Review*

*The Canadian Journal of Program Evaluation*, journal officiel de la Société canadienne d'évaluation

*Canadian Public Administration*

*Canadian Public Policy*

*Evaluation and Program Planning*

*Evaluation Practice*, anciennement *Evaluation Quarterly*

*Evaluation Review*

*Human Organization*

*International Review of Administrative Sciences*

*Journal of the American Statistical Association*

*Journal of Policy Analysis and Management*,

*Management Science*

*New Directions in Program Evaluation*, journal officiel de l'American Evaluation Association

*Optimum*

*Policy Sciences*

*Psychological Bulletin*

*Public Administration*

*Public Administration Review*

*The Public Interest*

*Public Policy*

*Survey Methodology Journal*

Le lecteur pourra aussi consulter avec profit d'autres publications sur l'évaluation portant sur des secteurs de programme particuliers, comme les services de santé, l'éducation, les services sociaux et la justice pénale.